

Declaration

Patterns of Evolution of Highly Repeated

The research carried

DNA from Caledia captiva

the results presented in this thesis have

acknowledged, the original work of

A thesis submitted for the degree of Doctor of Philosophy of The  
Australian National University.

Michael L. Arnold

July, 1985.

Michael L. Arnold

July, 1985.



## Table of Contents

Page

### Declaration

The research carried out in the course of this investigation and the results presented in this thesis are, except where acknowledged, the original work of the author.

Michael L. Arnold

*Michael L. Arnold*

July, 1985.



## Table of Contents

	Page
Acknowledgements . . . . .	i
Abstract. . . . .	ii
Chapter 1 - Introduction	
1.1 Literature Survey . . . . .	1
1.2 The Biology of Caledia. . . . .	17
Chapter 2 - Material and Methods	
2.1 Collection and Processing of Animals . . . . .	22
2.2 DNA Isolation. . . . .	22
2.3 Shearing of DNA and Isolation of a Highly Repeated Fraction. . . . .	24
2.4 Nick-Translation of DNA . . . . .	24
2.5 Synthesis of <sup>3</sup> H c-RNA for <u>in situ</u> Hybridization Probes. . . . .	28
2.6 Restriction Endonuclease Experiments . . . . .	29
2.7 Gel Electrophoresis. . . . .	29
2.8 Transfer of DNA from Agarose Gels to a Nylon Membrane. . . . .	31
2.9 Hybridization of <sup>32</sup> P-Nick-Translated DNA to DNA Fragments Immobilized on Gene Screen . . . . .	32
2.10 Electroelution of DNA from Agarose Gels. . . . .	34
2.11 Cloning <u>Caledia</u> Highly Repetitive DNA Using the Plasmid Vector pBR322. . . . .	35
2.12 Cloning <u>Caledia</u> Highly Repetitive DNA Using the Vector Phage 1059 . . . . .	40

2.13	Cloning of <u>Caledia</u> Highly Repeated DNA	
	into the Vector M13. . . . .	42
2.14	Melting Point (T <sub>m</sub> ) Determinations and	
	Saturation Hybridization Experiments. . . . .	45
2.15	<u>In situ</u> Hybridization of <sup>3</sup> H c-RNA from	
	<u>Caledia</u> Highly Repeated DNA to Prepared	
	Chromosomes . . . . .	46
2.16	DNA Sequencing . . . . .	48
2.16.1	Dideoxy Chain Termination Method. . . . .	48
2.16.2	Base-Specific Chemical Cleavage Method. . . . .	50
2.17	Isolation of Total RNA . . . . .	51
2.18	Electrophoresis of RNA Samples. . . . .	51
2.19	Transfer of RNA from an Agarose Gel to a	
	Nitrocellulose Filter and Hybridization	
	of Radioactive Probes to the Immobilized	
	RNA . . . . .	52
2.20	Computer Assisted Analysis of the Repeated	
	Sequences . . . . .	53

Chapter 3 - The Heterochromatin of Grasshoppers	
from the <u>Caledia captiva</u> species Complex I.	
Sequence Evolution and Conservation in a Highly	
Repeated DNA Family . . . . .	54

Chapter 4 - The Heterochromatin of Grasshoppers from	
the <u>Caledia captiva</u> species complex II. Cytological	
Organization of Tandemly Repeated DNA Sequences. . . . .	78

## Chapter 5 - Bam Family Repeated Sequences

5.1 Introduction . . . . . 103

5.2 Results and Discussion. . . . . 103

## Chapter 6 - Eco Repeated Sequences

6.1 Introduction . . . . . 110

6.2 Results and Discussion. . . . . 110

## Chapter 7 - Final Remarks

7.1 Introduction . . . . . 112

7.2 Evolution of Highly Repeated DNA Families . . 112

7.3 Functional Aspects of Highly Repeated  
Sequences . . . . . 117

Bibliography . . . . . 129



### Acknowledgements

I thank David Shaw and Rudi Appels for their guidance, many times in spite of my stubbornness. I also wish to thank them, along with Bernard John, for reading and vastly improving this thesis.

I profitted greatly from interactions with the following people:

R. Appels, A. Brown, I. Boussy, L. Christidis, D. Coates, D. Colgan, N. Contreras, S. Easteal, M. Fischer, R. Honeycutt, B. John, D. Llewellyn, L. McIntyre, G. Miklos, L. Moran, C. Moritz, D. Rowell, D. Shaw and P. Wilkinson.

I received excellent technical advice from R. Appels, B. Clarke, K. Ferguson, L. Moran, D. Shaw and P. Wilkinson.

K. Christian produced the computer program for comparing the repeated sequences.

The Commonwealth Scientific and Industrial Research Organization provided extensive support in the form of materials and equipment during the entirety of my tenure.

I would also like to acknowledge the encouragement I have received from my two sets of parents, Mr. and Mrs. R.C. Arnold and Mr. and Mrs. G.E. Swinson.

Finally, I dedicate this thesis to the two dearest people in my life, my wife Frances who has supported me in every way over the past ten years and my Lord Jesus Christ to whom I owe everything.

## Abstract

Five highly repeated DNA families were isolated from the grasshopper Caledia captiva. These families were analyzed with DNA reassociation, restriction endonucleases, DNA sequencing and in situ hybridization. Results from the first three analyses demonstrated that there was a large range of nucleotide sequence variation among repeats from these DNA families, from a low of 0% to a high of 16%. However, at least within the Taq family of repeats, there was a conserved region which showed no variation at the individual, populational, racial or subspecific levels. Furthermore, this conserved region formed a part of a region of dyad or two-fold symmetry, which is suggestive of protein-DNA interaction. The Taq sequence family was distributed in different chromosomal regions in the various taxa and, quantitatively, shows a negative correlation with the grade of reproductive isolation present in the species group. In contrast, the Bam family sequences are considerably more variable in nucleotide sequence than the Taq family sequences. Thus the tandem array of Bam sequences showed more interrepeat variability than the intersubspecific comparisons of the Taq repeats. This increased variation may be due to the age of the Bam family. Except for the comparisons of repeats from a tandem array, the Bam family sequences demonstrate equivalent amounts of variation at all taxonomic levels. The Bam family repeats apparently have similar copy numbers in each of the taxa. Three C. captiva

highly repeated Eco RI fragments from a single bacteriophage clone were also analyzed. In situ hybridizations indicated that these sequences were dispersed throughout the genomes of the C. captiva taxa, as well as C. species nova 1. Preliminary sequencing data suggested that these three sequences were related, but belong to different highly repeated DNA families. In addition, at least one of these repeats has a sequence structure which is necessary for forming the Z-DNA conformation. The results from this study of the Caledia repeated DNA sequences are discussed in terms of the evolution and potential functional significance of highly repeated sequence families.



## CHAPTER 1: INTRODUCTION

### 1.1 Literature Survey

An analysis of the organization of DNA within the genomes of eukaryotes was greatly facilitated by the discovery that, given the proper biochemical conditions, dissociated complementary strands of bacterial DNA would specifically reassociate (Marmur et al., 1963). Although this reassociation was defined by second-order kinetics, it was soon apparent that various portions of the eukaryote genome possessed differing "rates of reassociation" and differing levels of stability for the reassociated DNA (Britten and <sup>Kohne</sup>, 1968). These differences were subsequently attributed to dissimilarities in the repetition frequencies of various DNA sequences and the repetition frequencies were found to vary greatly within a genome. For example, the mouse genome was defined as consisting of approximately 70% unique or non-repetitive sequences, 20% sequences which are represented by  $10^3$ - $10^5$  copies and 10% sequences with  $10^6$  or more copies (Britten and <sup>Kohne</sup>, 1968). In the mouse genome, the latter fraction was defined as a highly repeated sequence "family" in the sense that sufficient homology was present between individual repeats which allowed reassociation. In contrast to the pattern of genome organization in the mouse, Flavell et al. (1974) found that the repeated sequences in 23 plant species constituted from 46-95% of the total genome. This type of measurement is dependent upon the conditions of reassociation (Moyzis et al., 1981). A more detailed examination of hexaploid wheat (Smith and Flavell, 1975), indicated that this genome was composed of a slow,

intermediate and rapidly reannealing fraction which constituted approximately 12%, 83% and 4-10% of the genome, respectively.

The initial analyses of highly repeated DNA families were reported almost a decade after the secondary structure of the DNA molecule had been defined by Watson and Crick (1953). A number of studies which utilized the technique of equilibrium ultracentrifugation of DNA in CsCl gradients, resolved multiple bands within a genome (Kit, 1961; 1962 a,b; Sueoka, 1961; Sueoka and Cheng, 1962; Szybalski, 1962). Sueoka (1961) detected a bimodal distribution in the guanine + cytosine content when selected samples of crab, calf and mouse DNA were subjected to this process. Likewise, Kit (1961; 1962 a,b) noted that the density gradients of mouse and guinea pig DNA differed from experiments involving both rat and human DNA in that the latter two demonstrated a single band following centrifugation. Although the presence of minor or "satellite" (Kit, 1961) bands in only two of three crab species examined led Sueoka (1961) to suggest that these bands might somehow be artifactual, both he and Kit (1962b) recognized that they could represent DNA which had a relatively higher ratio of adenine + thymine (A+T) to guanine + cytosine (G+C). This conclusion was drawn from the observation that an increase in the mean molar G+C content caused a linear increase in the mean buoyant density of DNA (Rolfe and Meselson, 1959). Therefore, these relatively less dense minor bands would represent DNA which had an overall lower G+C content. In fact, Flamm et al. (1967) proved this to be the case after determining the nucleotide composition of mouse satellite by

utilizing chromatography following enzymatic breakdown of the DNA into 5 - monophosphates.

Further studies of mouse satellite DNA specified that this component had a high degree of homogeneity both in terms of the molecular weights as well as the G+C content of the individual DNA molecules (Flamm et al., 1966). This homogeneity was indicated by both the sharpness of the mouse minor band formed in CsCl gradients and the three to four-fold narrower DNA melting curve for the satellite fraction relative to unfractionated mouse DNA. Moreover, it was noted that this DNA fraction had the ability to renature rapidly relative to other portions of the mouse nuclear and mitochondrial genomes and bacterial genomes (Bolton et al., 1965; Walker and McLaren, 1965; Flamm et al., 1966; Waring and Britten, 1966; Flamm et al., 1967). In light of the rate constant of reassociation calculated for this DNA, it was estimated that this fraction consisted of approximately  $1.0 \times 10^6$  tandemly repeated copies of a 300 - 400 base pair long sequence (Bolton et al., 1965; Waring and Britten, 1966).

In addition to the molecular characteristics of the satellite DNA, which suggested a structure composed of relatively simple, homogeneous components, the chromosomal location of rapidly reannealing DNA was defined cytologically in human, mouse and Drosophila genomes (Pardue and Gall, 1969; 1970; Jones, 1970; Rae, 1970; Jones and Corneo, 1971; Gall et al., 1971; Peacock et al., 1974). The human satellite II (Jones and Corneo, 1971) and mouse satellite (Pardue and Gall, 1969; 1970; Jones, 1970) were shown to be restricted to the centromeric regions within their respective genomes. Rae (1970) and Gall et al. (1971) found that



rapidly reannealing DNA components from Drosophila melanogaster, isolated both by hydroxyapatite chromatography and buoyant density gradients, showed a variety of hybridization sites including the chromocenter, the nucleolar organizer and a region near the tip of chromosome 3. Peacock et al. (1974) extended the analysis of the D. melanogaster highly repeated families by the isolation of previously undetected satellite sequences. This was accomplished by the utilization of antibiotic - CsCl gradients. The rationale for the use of such gradients was that certain antibiotics had been demonstrated to bind differentially to various DNA sequences, thus causing a change in their buoyant densities and thereby facilitating their separation in a CsCl gradient (Brown et al., 1971). Such satellites were later termed "cryptic" (see Singer, 1982 for review) because they were not detected in standard buoyant density gradients. In D. melanogaster, actinomycin-D-CsCl gradients resolved four major satellite peaks (Peacock et al., 1974). These satellites, as with those previously identified (Rae, 1970; Gall et al., 1971), demonstrated limited cytological distributions. The sites of hybridization of c-RNA's synthesized from the four satellites included the chromocenter, a single band in the 21 C-D region of chromosome 2 and the centromeric regions of each of the large autosomes, the X and chromosome 4 (Peacock et al., 1974). The localities of hybridization of the highly repeated fractions from each of these organisms, corresponded mostly to regions containing heterochromatin (Pardue and Gall, 1970; Jones and Corneo, 1971; Peacock et al., 1974).

On the basis of reassociation kinetics, repetitive sequences were shown to be ubiquitous in all but bacterial, viral and blue-green algae genomes (Britten and <sup>Kohne</sup>, 1967; 1968). However, repetitive DNA and in particular satellite sequences isolated from different organisms were found to be distinct from one another both on the basis of buoyant density and reassociation data (Hennig and Walker, 1970; Skinner and Kerr, 1971; Walker, 1971; Beattie and Skinner, 1972). Hennig and Walker (1970) found that, within and between the two rodent families Cricetidae and Muridae, variation existed in the banding position of minor satellites as well as in the relative amounts of these satellites. Even within a single genus (Apodemus) "remarkable variation" was detected between the satellites from four species (Hennig and Walker, 1970). Likewise, Skinner and Kerr (1971) discovered variation in the buoyant density values and relative amounts of light and heavy satellites from four species of crab and a single lobster species. In addition to the differences detected by buoyant density analyses, heterologous DNA-DNA hybridizations involving repeated sequences from a number of mammalian species revealed, on average, only 20% homology (Hoyer et al., 1964).

The unique properties of these sequences (ie. highly repeated, tandemly arranged and distinctive base composition between closely related species) led to proposals of novel evolutionary mechanisms to account for their formation and organization. Britten and <sup>Kohne</sup> (1967) and Southern (1970) suggested that such sequences had originated relatively recently as indicated by their species specificity. The former authors

also proposed that the repetitiveness of these families of sequences was the result of numerous copies of a sequence being made over a short time period by a process they referred to as saltatory replication. In addition, Southern (1970) suggested that evolution of the guinea pig  $\alpha$ -satellite could be explained by multiplication events followed by periods of no amplification, but with the occurrence of base pair mutations. This model was used to explain the differential repetition of sequence variants detected in the guinea-pig satellite via pyrimidine tract analysis (Southern, 1970). If a sequence variant arose early in the evolutionary history of the family and was subsequently amplified by unequal crossing-over it would have a relatively higher frequency than a variant which arose late in this process and, therefore, had not undergone the same level of amplification (Southern, 1970; Walker, 1971).

The view that highly repeated or satellite DNA was relatively simple in sequence, as suggested by earlier studies (Flamm et al., 1967), had to be modified when data from pyrimidine tract analysis (Southern, 1970; Peacock et al., 1974; Biro et al., 1975), quantitative estimates (Sueoka, 1961; Mazrimas and Hatch, 1972), restriction endonuclease mapping (Botchan, 1974; Southern, 1975a) and DNA sequencing (Fry et al., 1973; Brutlag and Peacock, 1975; Endow et al., 1975; Shmookler Reis and Biro, 1978) were considered. All of these suggested the presence of a highly complex and integrated organization for these sequence families. For example, whilst the D. melanogaster highly repeated DNA families demonstrated intra-satellite



sequence variation (Peacock et al., 1974; Brutlag et al., 1977), each of the four major satellites had related sequences (Brutlag and Peacock, 1975). Strachan et al. (1982) using, in particular, the techniques of restriction endonuclease mapping and filter hybridization also detected interspecific homologies for Drosophila satellites as well as sequence homologies between some of these satellites. In addition, these authors detected variation in the repeat lengths of the "500 bp" satellite sequences using the restriction enzyme analysis. More recent analyses of satellite sequences from other animal taxa, involving nucleotide sequencing of cloned repeats (bovine: Plucienniczak et al., 1982; Tapasowsky and Gerbi, 1982; nematode: Muller et al., 1982; mouse: Pietras et al., 1983; african clawed-toad: Meyerhoff et al., 1983; sheep: Novak, 1984; and rat: Epstein et al., 1984) have also demonstrated the presence of i) sequence heterogeneity within highly repeated DNA families and ii) homology among satellites from a given species.

The utility of nucleotide sequence data, in terms of defining the organization of a highly repeated DNA family, is discussed here using the studies of the primate " $\alpha$ -satellite" sequence (Maio, 1971) as an example. The "fundamental" unit of the  $\alpha$ -satellite DNA was determined to be a 170 bp repeat as measured by restriction endonuclease cleavage. However, various primate species have repeat sequences which are defined as multiples (ie.  $170n$ ;  $n=1,2$  or  $4$ ) of the fundamental unit (Donehower and Gillespie, 1979). Thus, baboons (Papio), macaques (Macaca) and mangbeys (Cercocebus) are distinguished by a basic repeat which is 340 bp long. Guenon (Cercopithecus) species, on the

other hand, have a 170 bp repeat, whilst the colobus monkey (Colobus) has a repeat unit of 680 bp. The initial sequence analysis of the primate  $\alpha$ -satellite was carried out on the African Green monkey (AGM) repeat family (Rosenberg et al., 1978). This analysis demonstrated that the AGM repeats were not identical, but rather constituted a family of closely related 172 bp sequences. Rosenberg et al. (1978) showed that the repeat units were not internally repetitive and that divergence at the 172 bp positions was not random, but occurred more frequently at some positions than at others. In a further analysis of the AGM  $\alpha$ -satellite, Thayer et al. (1981) found that neighboring repeats were no more similar to one another than randomly isolated 172 bp units. Sequence analysis of human  $\alpha$ -satellite identified a 171 bp monomer unit which was 65% homologous to the 172 bp AGM repeat (Manuelidis and Wu, 1978). Both the human and the AGM sequences were shown to have numerous stop codons as well as several regions of dyad symmetry (Manuelidis and Wu, 1978). An analysis of  $\alpha$ -satellite repeats from the baboon defined a 343 bp sequence (Donehower et al., 1980). This repeat unit was shown to consist of two "wings" of 172 bp and 171 bp which shared 69% homology. In contrast, the 172 bp wing was 92% homologous to the AGM repeat unit. A comparison of the baboon, AGM and human  $\alpha$ -satellites led to the suggestion that each of their respective repeats had a common ancestor (Donehower et al., 1980). An examination of  $\alpha$ -satellite DNA from the bonnet monkey (Macaca radiata) genome revealed a similar pattern of organization to that found in the baboon (Rubin et al., 1980). The basic repeat in the bonnet

monkey was shown to be a 344 bp sequence which consisted of two related 172 bp sequences (Monomer I and II). Results from the comparison of these two sequences to each other as well as to the human and AGM  $\alpha$ -satellite DNA are shown in Table 1-1 (derived from Rubin et al., 1980). Both of the 172 bp sequences demonstrated greater homology with the AGM sequence than they did with each other. These two monomer units shared less homology with the human repeat (Table 1-1). Recent analyses of the  $\alpha$ -satellite sequences have involved the detailed examination of junctions between these sequences and non- $\alpha$ -satellite DNA (McCutchan et al., 1982; Maresca and Singer, 1983; Maresca et al., 1984). The AGM satellites at such junctions were found to be not only divergent forms of the major  $\alpha$ -satellite sequence (McCutchan et al., 1982), but in some instances novel satellites were identified (Maresca and Singer, 1983; Maresca et al., 1984).

The D. melanogaster satellites were shown to be distributed in numerous chromosomal locations such that each chromosome had "... a quantitatively and qualitatively unique segmented makeup of its heterochromatin" (Peacock et al., 1978). This segmental pattern of organization had previously been documented in the red-necked wallaby (Dunsmuir, 1976) and was also demonstrable for satellites from various plant (cereal) genomes (Peacock et al., 1978; Appels and Peacock, 1978).

As noted above, the complex pattern of organization between highly repeated DNA families included sequence relatedness among satellites from a given species. Furthermore, comparative analyses have also detected sequence homologies between the highly repeated fractions from related species (Sutton and



Table 1-1. Percent homology between the  $\alpha$ -satellite monomers from the bonnet monkey and between these sequences and the  $\alpha$ -satellite DNA from human and african green monkey (AGM) genomes. This table was derived from Rubin et al. (1980).

	<u>HUMAN</u>	<u>AGM</u>	<u>MONOMER I</u>
--	--------------	------------	------------------

Monomer I	67%	90%	100%
-----------	-----	-----	------

Monomer II	56%	72%	66%
------------	-----	-----	-----

McCallum, 1972; Gall et al., 1974; Peacock et al., 1978).

However, these sequence homologies were in contrast to the marked quantitative differences for a given satellite present in related species (Peacock et al., 1978). The obvious question generated from these findings was: How could a repeated sequence family which possesses up to  $10^6$  copies remain relatively homogeneous in terms of its basic sequence and yet fluctuate dramatically in quantity between related species? As mentioned previously a number of mechanisms had been envisaged to explain the amplification and evolution of these sequences (Britten and Kohne, 1967; Southern, 1970). Smith (1976) demonstrated mathematically that the process of unequal crossing over will, by chance, lead to the "homogenization" of an originally heterogeneous array of tandem repeats. This homogenization results in the predominance of one sequence variant. In addition, the simulations utilized to define the mechanism of unequal crossing over also suggested that in the absence of selective forces, the "natural state" of a DNA sequence was that of a tandem array (Smith, 1976). Subsequent to the analysis by Smith (1976), studies of yeast ribosomal RNA loci have indicated that unequal crossing over can occur between sister chromatids during both mitosis (Szostak and Wu, 1980) and meiosis (Petes, 1980).

In addition to the amplification and "homogenization" of a sequence array, it is also apparent that highly repeated DNA families have in some cases spread from a single to many chromosomal locations. Botchan (1974) suggested that a "rolling



circle" model of DNA amplification was the most plausible explanation for the pattern of organization observed in the bovine satellite I DNA. An inherent component of this process is the formation of rings of satellite DNA which are excised products of reciprocal intrastrand recombination. It is proposed that these rings then undergo a rolling circle mechanism of amplification analogous to that exhibited by rDNA in Xenopus oocytes (Hourcade et al., 1973). This results in an array of identical tandem repeats which have been postulated to reintegrate into the germ line genome (Botchan, 1974). If this reintegration occurs in a nonhomologous chromosomal region the sequence would then be established in a novel position. More recently, the discovery of moderately repetitive elements which have the ability to transpose either themselves or, in the case of the prokaryotic IS elements, entire chromosome segments from one site to another (see Iida et al., 1983 for review) has suggested a further mechanism capable of dispersing highly repeated DNA sequences throughout the genome. Although it is speculative, a transposition event involving a highly repeated DNA from a tandem array could involve a process whereby the repeated sequence was positioned between two such transposable elements and subsequently shifted along with the transposons to a novel chromosomal location.

In the above discussion the question was raised concerning how a sequence family, which might possess as many as  $10^6$  copies, could remain relatively homogeneous. The data which suggest homogeneity within a sequence family will be discussed in detail in the concluding chapter. Initially this homogeneity may very

well be facilitated by the mechanism of amplification (eg. rolling circle model). Maintenance of homogeneity could also be accounted for by unequal crossing over in instances where the sequence family is localized on a single pair of homologous chromosomes (Smith, 1976). However, if a sequence family is present on nonhomologous chromosomes, as for example in the mouse satellite (Pardue and Gall, 1970), factors other than recency of origin may account for the sequence similarity among the repeats. One such mechanism is gene conversion. In its classical sense, gene conversion occurs when one allele at a locus converts a second allele at an homologous locus into its own allelic state (Roman, 1963). Such intragenic conversion results in an atypical ratio of the segregants produced by meiosis (Case and Giles, 1964; Hurst et al., 1972). A similar process has seemingly been responsible for intergenic conversion events in the chorion and globin multigene families (Iatrou et al., 1984; Slightom et al., 1980) and in a duplicated leucine gene (Klein and Petes, 1981). For a new sequence variant to spread throughout a family of repeats, as the result of conversion events, the conversion must be "biased" (ie., it must favor the new variant). This type of biased conversion has been observed for intragenic recombination involving spore color loci in Ascobolus immersus (Leblon, 1972) and Sordaria brevicollis (Yu-Sun et al., 1977).

Mechanisms which may account for the generation, amplification, spread and maintenance of highly repeated DNA families have been described in the above discussion. However, one is still left with the question of whether or not highly

repeated DNA plays any functional role or roles within the genome.

The failure to detect overt cases of specific highly repeated DNAs with specific functions, has led various authors to suggest that these DNA families may be selfish (Doolittle and Sapienza, 1980) or parasitic (Orgel and Crick, 1980) in terms of their replication, neutral in terms of phenotypic expression (Doolittle and Sapienza, 1980; Orgel and Crick, 1980; Miklos and Gill, 1981), or the products of a process which selects and amplifies non-adaptive variants and spreads them between different chromosomes (Dover, 1982; Ohta and Dover, 1984). Each of these hypotheses regards highly repeated sequences as functionless although it has been suggested that the gradual accumulation of non-adaptive variants via the process of "molecular drive" (Dover, 1982) might result in reproductive isolation (Ohta and Dover, 1984). In contrast, the distinctive characteristics of satellite DNA, themselves, have led many investigators to conclude that this highly repeated fraction had a specific functional aspect. Thus Walker (1971), after noting that satellite DNAs appeared to be unique between closely related species and that they could vary extensively in quantity, nevertheless, concluded that "... the presence of satellite DNA confers some selective advantage; otherwise we cannot account for its spread throughout a population as large as that of the house mouse." Walker (1971) suggested specifically that the selective advantage might occur at the level of the chromosome, in that a homologue which possessed satellite sequences might have a higher chance of surviving the "mechanical processes" of meiosis than



its homologue which lacks such a sequence. Similarly, Mayfield and Ellison (1975) suggested that satellite DNAs might provide sites of recognition for homologous chromosome pairing at meiosis. Highly repeated DNAs have also been cited as possible candidates for acting as sterility barriers between species (Corneo, 1976; Fry and Salser, 1977) and in facilitating high genetic flexibility (Mazrimas and Hatch, 1972). Both Corneo (1976) and Fry and Salser (1977) hypothesized that the putative sterility barrier caused by differences in satellite DNA composition could be due to disruption of meiotic pairing sites associated with the highly repeated DNA. It should be noted, however, that with respect to each of the above hypotheses, John and Miklos (1979) concluded that none could be supported by experimental data.

A further factor which has intimated that highly repeated DNA may have functional attributes was the discovery of sequence homologies between both closely and more distantly related taxa. In spite of the numerous examples of such sequence conservation (Gall et al., 1974; Fry and Salser, 1977; Peacock et al., 1977; Peacock et al., 1978; Manuelidis and Wu, 1978; Witney and Furano, 1983; Pietras et al., 1983; Epstein et al., 1984; Dennis and Peacock, 1984; Novak, 1984), Miklos (1985) has pointed out that without estimates of species variation "... hypotheses relating to conservation (of highly repeated sequences) are severely constrained."

By combining molecular and cytogenetic analyses, Miklos and Nankivell (1976) established that heterochromatic blocks in the

grasshopper Atractamorpha similis possessed a large amount of satellite-1 sequences which had an effect on the pattern of male meiotic recombination. Specifically, in chromosomally polymorphic populations of Atractamorpha, the presence of large telomeric blocks of satellite DNA resulted in a redistribution of chiasmata away from the blocks (Miklos and Nankivell, 1976). This type of effect has also been documented in a variety of other grasshopper species both for telomeric and centromeric heterochromatin (John and King, 1985). It is important to note, however, that this effect does not appear to be present in the chromosomal race of the grasshopper Caledia captiva which possesses large centromeric blocks of highly repeated DNA (Shaw and Knowles, 1976; Appels and Peacock, 1978). A second attribute which has been demonstrated for several highly repeated sequences, and which may have functional significance, is the provision of specific sites for DNA-protein binding (Hsieh and Brutlag, 1979a; Levinger and Varshavsky, 1982 a, b; Avila et al., 1983; Strauss and Varshavsky, 1984; Andrews et al., 1984). Data from these studies suggest that specific functions for the various protein - DNA interactions may include nucleosome phasing (Strauss and Varshavsky, 1984), involvement in microtubule - chromosome association (Avila et al., 1983) and the influencing of gene activity (e.g., Andrews et al., 1984).

The above discussion clearly indicates the immense complexity of highly repeated DNA in terms of its origin, maintenance, evolution and potential for functional interactions. Some of these aspects have been examined with the techniques of recombinant DNA technology. However, no study to date has

examined a group of related taxa with the view of understanding how highly repeated sequences originate, evolve and interact in their respective genomes. The research work described in this thesis is directed toward resolving the following questions:

- 1) What is the level and distribution of sequence variation in specific highly repeated DNA families within an individual animal and a population of animals compared to that present between different populations of the same species which show increasing degrees of reproductive isolation?
- 2) How are the various sequence families organized at the chromosomal level?
- 3) What, if any, is the relationship between the families in terms of their sequence composition?
- 4) Do the patterns of evolution in these highly repeated DNAs agree with models which predict how such DNA should evolve?
- 5) Are the data from this study of specific highly repeated DNA families suggestive of any function?

In order to approach the above questions, there are specific attributes which must be present in the organism chosen for analysis. First, the selected species must contain taxa which demonstrate a grade of reproductive isolation from local populational status though to taxonomic species. Second, the chromosomes which belong to the various taxa must be individually and unambiguously distinctive, possessing regions of cytologically definable heterochromatin. Finally, the chromosome



complements from the various taxa must demonstrate intra- and intertaxonomic variation in their patterns of heterochromatin distribution. As will be documented in Section 1.2, grasshoppers of the genus Caledia captiva provide all of these biological attributes.

## 1.2 The Biology of Caledia

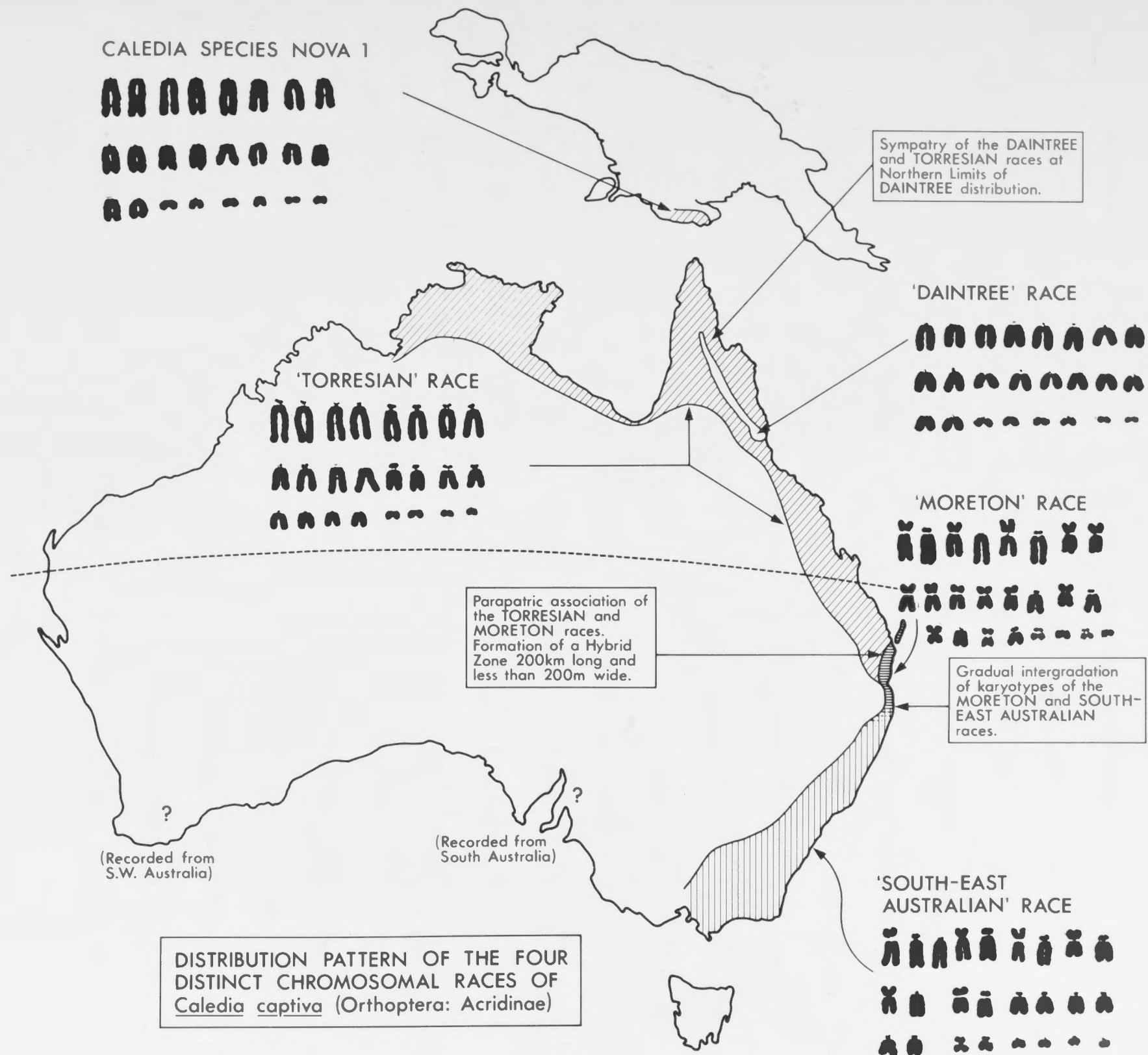
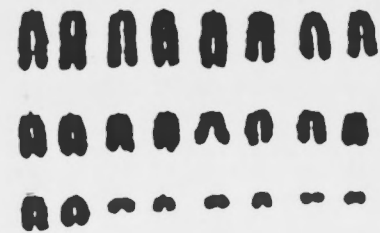
The organism chosen for this study was the grasshopper genus Caledia (Orthoptera: Acrididae: Acridinae). Individuals from both of the conventionally recognized taxonomic species, Caledia captiva and Caledia species nova 1, were analyzed. This genus is currently known from localities ranging from 8° - 38° latitude, and so includes within its distributional range, populations in Papua New Guinea, the Torres Strait region, northern coastal Australia and the entire eastern seaboard of Australia. Specimens of C. species nova 1, which is restricted to Papua New Guinea, have been collected from monsoon rain forest on the Oriomo Plateau (Fig. 1-1). In contrast, C. captiva is found throughout the entire distributional range of the genus and so occurs in a variety of habitats including tropical rain forest, tall mixed savannah, open forest and grassy open forest (Fig. 1-1). C. captiva and C. species nova 1 are easily distinguishable on the basis of both size and coloration (Shaw, 1976).

Both species possess a karyotype of 11 autosomal pairs of chromosomes and an XX(♀)/X0(♂) sex chromosome system. This results in 23 chromosomes in the male and 24 chromosomes in the female. The C. species nova 1 karyotype is monomorphic, in terms

Figure 1-1. Geographic distribution of the Caledia taxa.

Note: The Torresian taxon also occurs sympatrically with C.  
species nova 1. This figure was taken from Shaw et al. (1980).

CALEDIA SPECIES NOVA 1





of centromere position (Shaw, 1976), and all chromosomes are telocentric (Fig. 1-2). C. captiva, on the other hand, shows extreme levels of chromosomal variation which allows the recognition of four chromosomal taxa (Shaw, 1976; Shaw et al., 1976; Shaw et al., 1980). The Daintree (DT) taxon has a telocentric complement (Table 1-2; Fig. 1-2), the Torresian (TT) taxon possesses an acrocentric karyotype, (Table 1-2; Fig. 1-2), the Moreton (MM) taxon exhibits a metacentric complement (Table 1-2; Fig. 1-2) and the South-east Australian (S.E.A.) chromosomal form is characterized by a mixture of metacentric, acrocentric and telocentric chromosomes (Table 1-2; Fig. 1-2). The complexity of the chromosomal system of C. captiva is further evidenced when techniques are utilized to identify C-band heterochromatin (Shaw et al., 1976; Shaw et al., 1980; Coates and Shaw, 1984; Shaw et al., in prep.). C-band variation is extensive both within and between populations. Hence, Daintree populations are characterized by large procentric blocks of heterochromatin (Table 1-2; Fig. 1-2), Torresian populations by centromeric or short arm heterochromatin (Table 1-2; Fig. 1-2) and Moreton and S.E.A. populations by numerous interstitial C-band positive regions (Table 1-2; Fig. 1-2). In addition, C. species nova 1 individuals possess mainly small procentric regions of C-band heterochromatin (Fig. 1-2).

Additionally, there is a cline in centromere position between the Moreton and S.E.A. populations (Shaw et al., 1980; Shaw et al., in prep.). The limits of this cline are represented by a homozygous metacentric karyotype in the northern Moreton populations and a totally acrocentric karyotype in the southern

TABLE 1-2. Karyotypic attributes of the C. *captiva* chromosomal  
taxa and C. *species nova* 1.

<u>SPECIES</u>	<u>RACE</u>	<u>CENTROMERE LOCATION</u>	<u>HETEROCHROMATIN DISTRIBUTION</u>
<u>C. captiva</u>	Daintree	Telocentric	Large Procentric Blocks
<u>C. captiva</u>	Torresian	Acrocentric	Centromeric and Short Arms
<u>C. captiva</u>	Moreton	Metacentric	Interstitial (Polymorphic)
<u>C. captiva</u>	South-east Australian	Polymorphic	Interstitial (Polymorphic)
<u>C. species nova 1</u>	---	Telocentric	Small Procentric Blocks



**Figure 1-2.** C-banded karyotypes of the Caledia taxa.

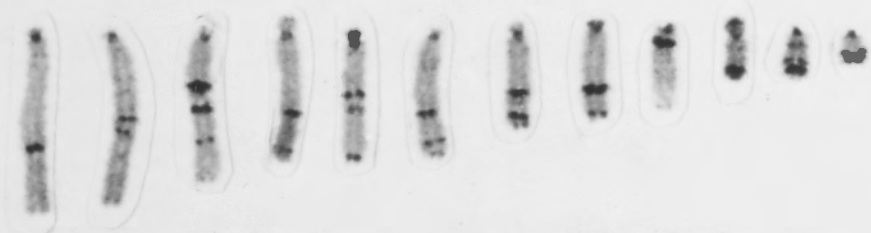
Note: Southern, Northern and Papuan Torresian are differentiated within the Torresian taxon on the basis of cytotype (Southern vs Northern) or geographic distribution (Papuan).

1 2 X 4 5 6 7 8 9 10 11 12

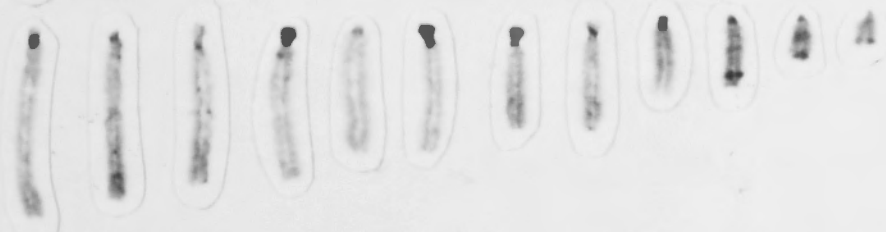
*C. captiva*  
(MORETON)



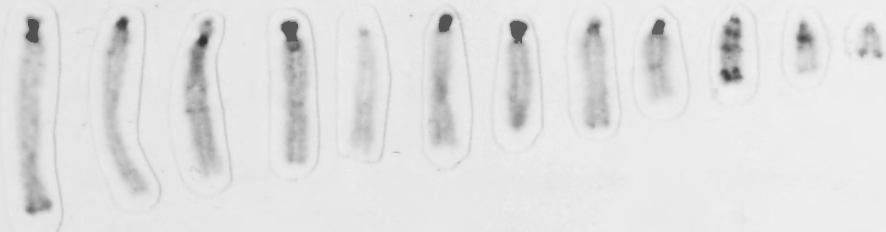
*C. captiva*  
(LAKES  
ENTRANCE)



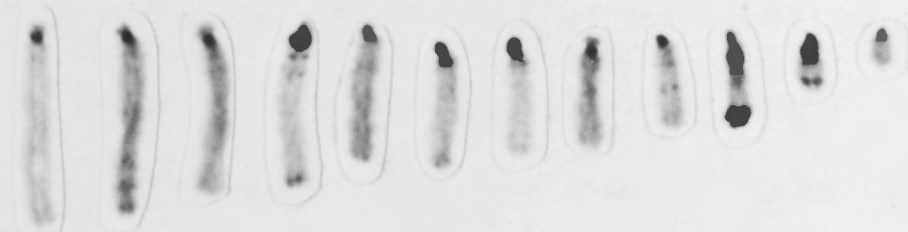
*C. captiva*  
(SOUTHERN  
TORRESIAN)



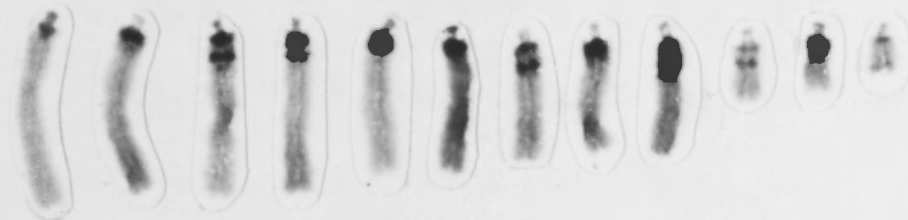
*C. captiva*  
(PAPUAN)



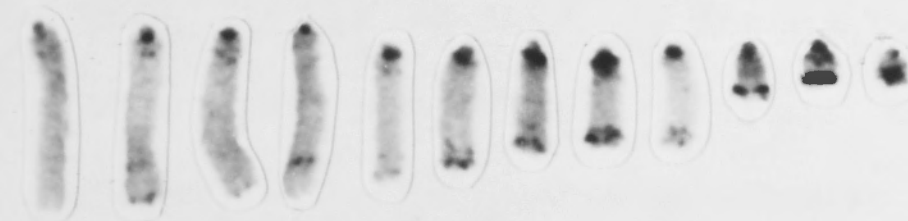
*C. captiva*  
(NORTHERN  
TORRESIAN)



*C. captiva*  
(DAINTREE)



*C. species*  
nova 1



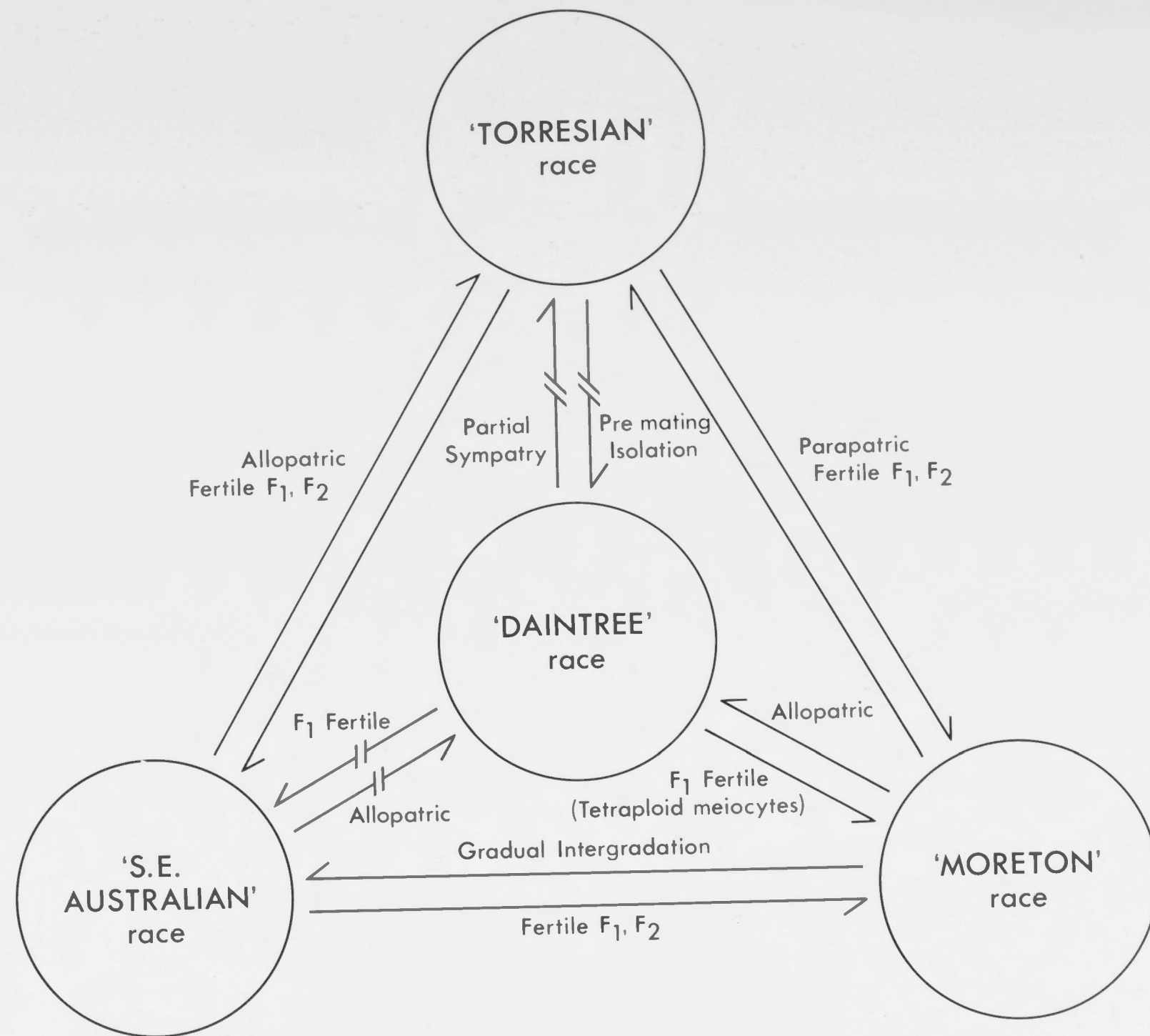
S.E.A. populations (Shaw and Coates, 1983). It is of interest that both of these populations from the ends of the cline still maintain the interstitial C-band heterochromatin, as do the intermediate populations (Shaw et al., in prep.).

Laboratory crosses, along with field observations, indicate that differing levels of reproductive isolation are present among the various Caledia taxa (Shaw et al., 1980; Coates and Shaw, 1984; Fig. 1-3). The Daintree form is reproductively isolated from each of the other three taxa by premating mechanisms in the case of the Torresian chromosomal type and both pre- and postmating in respect of both the Moreton and S.E.A. taxa. Although the Daintree and Torresian taxa occur sympatrically in at least some populations (Shaw et al., 1980), they do not appear to form natural hybrids (Shaw, pers. comm.). Cytogenetic analysis of meiosis in F1 hybrids from (MM X DT) and (S.E.A. X DT) crosses have detected the presence of numerous aberrant configurations involving "nonhomologous" associations between chromosomes from the same genome (Shaw and Wilkinson, 1978). The Torresian form, when crossed to either the S.E.A. or Moreton race, produces fertile F1's. However, in the (MM X TT) crosses there is complete F2 breakdown while the (S.E.A. X TT) matings produce from 25 - 54% viable F2's (Daly et al., 1981; Coates and Shaw, 1984). In the (MM X S.E.A.) crosses the resulting F1 hybrids are fully viable (Shaw et al., 1980; Shaw et al., in prep.). However, in the F2 generation there is a range of viabilities (58-100%) depending upon which S.E.A. population is utilized (Shaw et al., 1980; Shaw and Coates, 1983).



Figure 1-3. Schematic representation of the reproductive isolation present between the C. captiva taxa.

Note: This figure was taken from Shaw et al. (1980).



In relation to the above findings, Daly et al. (1981) in a study of the pattern of allozyme divergence among the C. captiva taxa, established that genetic distances (Nei and Roychoudhury, 1974) increased directly with the level of reproductive isolation. However, these authors concluded that the divergence in the genic component within animals from the various taxa had not given rise to the reproductive isolation seen in the present population (Daly et al., 1981). Data which support this hypothesis are available from crosses involving the Lakes Entrance S.E.A. population and the Moreton population. These two populations are equivalent in terms of their genic component (Daly et al., 1981) yet when crossed the F2 generation shows a 42% reduction in viability. This is directly attributable to chromosomal differences between the two populations (Shaw and Coates, 1983).

A more precise definition of some of the elements responsible for the reproductive isolation of the Moreton and Torresian taxa, has been facilitated by the discovery and subsequent analysis of a zone of hybridization between them (Moran and Shaw, 1977; Moran, 1979; Shaw et al., 1979; Shaw and Wilkinson, 1980; Moran et al., 1980; Shaw et al., 1985). This zone is situated in South-east Queensland and extends latitudinally for at least 150 km (Moran and Shaw, 1977). In spite of the extensive parapatric association in this region, the changeover in the frequency of eight of the twelve members of the Torresian and Moreton genomes occurs over a distance of only 200 meters (Moran, 1979). It was originally demonstrated that certain of the Torresian chromosomes were introgressing into the



Moreton subspecies, but without reciprocal introgression (Moran and Shaw, 1977). Earlier work on the chromosomal parameters of the C. captiva complex, as well as the hybrid zone data, led Moran and Shaw (1977) to conclude that this contact zone was "... probably a result of an interaction between both ecological tolerance and competitive exclusion". The maintenance of the abrupt changeover in the frequencies of Torresian and Moreton chromosomes was attributed to total embryonic breakdown in the F2 generation and approximately 50% backcross inviability (Shaw et al., 1980). This was initially hypothesized, and subsequently demonstrated, to be due to the redistribution of chiasmata into chromosome regions which do not normally undergo recombination (Shaw and Wilkinson, 1980; Coates and Shaw, 1982; 1984; Shaw et al., 1982). It has been suggested that this pattern of novel recombination perturbs the internally coadapted gene complexes present within the chromosomes of a given complement thereby leading to dysfunctional combinations (Shaw and Coates, 1983). More recently it has been demonstrated that the structure of the hybrid zone is also due in part to the action of selection which favors the metacentric Moreton karyotype during mesic years and the acrocentric Torresian karyotype in arid years (Shaw et al., 1985). These data have led to the proposal that the overall structure of the karyotype may have a functional significance over and above its genic content (Shaw et al., 1985; Shaw et al., in prep.).

Thus, Caledia provides an ideal system in which to examine each of the questions posed in Section 1.1.

## CHAPTER 2: MATERIAL AND METHODS

### 2.1 Collection and Processing of Animals

The majority of the animals used in this research project were collected from wild populations. The remainder of the samples were obtained from offspring of the wild caught animals. Within the laboratory, grasshoppers were maintained in 45cm X 37cm X 37cm cages with a 75 watt photoflow incandescent light source. These cages were kept at 26-28°C and 50-60% relative humidity during the daytime (14 hr) and 16-18°C and 70-85% relative humidity during the night time (10 hr). A mixture of Rhodes (Chloris gayana) and Paspalum grass, wheat seedlings and rabbit pellets, as well as a moistened pad of cotton wool, were maintained in each of the cages. In addition, plastic containers (10cm diameter X 12cm depth) filled with moist sand were kept in each cage to facilitate the laying of eggs. These containers were checked on alternate days for the presence of egg pods. Egg pods were kept at 30°C until the embryos were needed for experimentation or until hatching occurred ( $28 \pm 1$  days). Prior to their use, adult grasshoppers were killed, eviscerated and stored in liquid nitrogen.

### 2.2 DNA Isolation

DNA, from either pooled samples or individual grasshoppers, was isolated using a modification of the protocol of Appels and Dvorak (1982). Whole, eviscerated animals were added to a mortar along with approximately 1 gm of acid washed sand (BDH). Liquid nitrogen was then added and the immersed grasshopper(s) ground with a pestle to a fine powder. Additional liquid nitrogen was

added as necessary to ensure that the sample was immersed. The powder obtained was transferred to a 15 ml corex centrifuge tube and 1 ml of buffer (0.1M NaCl, 0.1M EDTA, 0.05M Tris-Sigma, pH 7.5), 100  $\mu$ l of 5% Lauryl Sulphate (SDS-Sigma) and 100  $\mu$ l of Proteinase K (Boehringer - 0.2 mg/ml) were added. Following vigorous agitation of this mixture, the tube was covered with parafilm and incubated at 37°C for from 30 min to 3 hr while other samples were being processed. One ml of phenol was then added to each tube and the samples agitated occasionally at room temperature prior to centrifuging at 8000 rpm for 5 min. The supernatant was transferred to a second 15 ml corex tube, using a cut pasteur pipette to avoid shearing the DNA. One ml of a Tris-EDTA (TE) buffer (0.01M Tris, 0.001M EDTA, pH 8.4) was added to the original tube, the tube was agitated, respun and the supernatant was transferred to the previously collected sample. In order to precipitate the native DNA from the supernatant, an amount equivalent to 0.1 of the supernatant volume of sodium acetate (2M, pH 5.5) and two volumes of absolute ethanol were added. Upon addition of these solutions, the DNA became visible as long strands. The DNA was collected either by "spooling" via a pasteur pipette or by centrifuging the sample at 10,000 rpm for 5 min. The DNA was then added to 1 ml of TE buffer and 0.2 ml of 20 X SSC (3M NaCl, 0.3M NaCitrate), and after it had dissolved, 1.2  $\mu$ l of RNA'ase A (10 mg/ml Boehringer) was added and the solution was incubated at 37°C for 30 min. Fifty  $\mu$ l of 5% SDS and 1 ml of phenol were added and the mixture was treated as described above. Subsequent to the ethanol precipitation, the



DNA was dissolved in 1 ml of TE and stored at  $-20^{\circ}\text{C}$  until its use.

### 2.3 Shearing of DNA and Isolation of a Highly Repeated Fraction

Isolated DNA which had been stored in the TE buffer was dialyzed into a 0.12M sodium phosphate buffer. This DNA was then passed three times through a french press at 16,000 psi in order to shear the DNA to an average size of approximately 500 bp. The sheared DNA was heat denatured and allowed to renature to a Cot value of approximately 0.02. This value represents a point on the Cot curve which is in the range of highly repeated DNA (Britten and <sup>Kohne</sup>, 1967). The entire DNA sample was then loaded onto a hydroxylapatite column using a continuous column loading system (Appels et al., 1978). After loading, the column was initially rinsed with aliquots of the 0.12M sodium phosphate buffer to ensure that single-stranded DNA was washed from the hydroxylapatite. Subsequently, 5 ml aliquots of 0.5M sodium phosphate buffer were passed across the column in order to elute the double-stranded molecules. This fraction was collected, dialyzed into TE buffer and stored at  $-20^{\circ}\text{C}$ .

### 2.4 Nick-Translation of DNA

Radioactive probes used in various hybridization experiments were produced using the nick-translation method of Rigby et al. (1977). For each nick-translation reaction, 0.5 - 1  $\mu\text{g}$  of the appropriate DNA was used. Following ethanol precipitation and vacuum drying of the DNA, the sample was resuspended in 37  $\mu\text{l}$  of  $\text{H}_2\text{O}^1$ . The following solutions were added to an acid washed tube;

5  $\mu$ l of [ $\alpha$ - $^{32}$ P]dATP (2'-deoxyadenosine 5'-triphosphate, 5 $\mu$ Ci/ $\mu$ l, Amersham), and 15  $\mu$ l each of 0.5mM dCTP (2'-deoxycytidine 5'-triphosphate, P-L Biochem.), dGTP (2'-deoxyguanosine 5'-triphosphate, P-L Biochem.) and dTTP (thymidine 5'-triphosphate, P-L Biochem.). This solution was then vacuum dried using a rotary vacuum and was subsequently resuspended in 8  $\mu$ l of the cocktail mix (130  $\mu$ l of 0.25M NaPO<sub>4</sub>, 5  $\mu$ l of 1M MgCl<sub>2</sub>). The reaction mixture was added to a 1.5 ml eppendorf centrifuge tube and consisted of the DNA sample (37  $\mu$ l), 6  $\mu$ l of the triphosphate solution, 1  $\mu$ l of DNA'ase (0.01 mg/ml, Boehringer), 1  $\mu$ l of 50 $\mu$ M dATP (2'-deoxyadenosine 5'-triphosphate, P-L Biochem) and 5  $\mu$ l of DNA polymerase I (Kornberg polymer, Biolabs). This solution was incubated at 14°C for 1 hr. To monitor the extent of the reaction, 1  $\mu$ l of the reaction mixture was then added to 100  $\mu$ l of H<sub>2</sub>O. Ten microlitres of this dilution was added to 50  $\mu$ l of sonicated salmon sperm DNA (2 mg/ml) and 1 ml of 5% trichloroacetic acid. This mixture was placed on ice for 5 min to allow precipitation and was then loaded onto a glass fibre filter (Whatman), washed twice with 2 ml aliquots of 0.1M Na<sub>2</sub>H<sub>2</sub>P<sub>2</sub>O<sub>7</sub> (Sigma) in 1M HCl and finally with approximately 10 ml of absolute ethanol. This filter represented the incorporated radioactivity. The total amount of radioactivity was measured by spotting 10  $\mu$ l of the diluted reaction sample onto a glass fibre filter. Replicate filters were prepared for both the total and incorporated measurements. The filters were dried, placed in

<sup>1</sup> Unless otherwise noted, the H<sub>2</sub>O used in the experiments was distilled and then sterilized by autoclaving.

scintillation vials, scintillant (0.06% 2,5-Diphenyloxazole, Calbiochem and 0.006% 1,4-di[2-(5-Phenyloxazolyl)]-benzene Calbiochem, in Toluene) was added and the radioactivity present in each vial measured using a Packard Tri-Carb Liquid Scintillation Spectrometer. The percentage of incorporation of radioactive nucleotides was then calculated. In order to remove the enzymes present in the reaction mixture, 100  $\mu$ l of TE, 25  $\mu$ l of 0.25M EDTA, 5  $\mu$ l of 5% SDS, 100  $\mu$ l of phenol and 100  $\mu$ l of chloroform were added. Following vigorous vortexing, the phenol/chloroform and aqueous phases were separated by centrifugation for 1 min in an Eppendorf (Model 5414) centrifuge. The aqueous phase was transferred with a pipettman to a fresh eppendorf centrifuge tube. The phenol/chloroform was then washed with an additional 50  $\mu$ l of TE and, following separation of the phases, the aqueous portion was added to the previously removed aliquot. This sample was ethanol precipitated after which the pellet was washed with 70% ethanol and then vacuum dried. In order to increase the single-strand length of the nick-translated DNA from 50-100 nucleotides up to approximately 300 nucleotides (Appels and Dvorak, 1982), the sample was ligated using T4 DNA ligase (see Higgins and Cozzarelli, 1979 for a review of the properties of T4 ligase). The pellet was resuspended in 13  $\mu$ l of H<sub>2</sub>O, 4  $\mu$ l of 5 X ligase buffer (50mM Tris-HCl pH 7.8, 10mM MgCl<sub>2</sub>, 50  $\mu$ g/ml bovine serum albumin), 2  $\mu$ l of 10 mM ATP, 0.2  $\mu$ l of 1M DL-Dithiothreitol (DTT-Sigma) and 0.5  $\mu$ l of T4 DNA ligase (4 X 10<sup>5</sup> units/ml-Biolabs). This reaction was incubated at 15°C from 2 - 20 hrs. Subsequently, 1  $\mu$ l of 5% SDS and 6  $\mu$ l of proteinase K (0.1 mg/ml) were added and the mixture was incubated at 37°C



for 30 min. The sample was then extracted (as described above) with 100  $\mu$ l TE, 130  $\mu$ l phenol and 130  $\mu$ l chloroform and the phenol/chloroform phase was washed with an additional 100  $\mu$ l of TE. The aqueous portion was loaded onto a sephadex G-75 column (8mm X 135mm) (Pharmacia) and 20 X 14 drop samples were collected using a Gilson Micro Fractionator. The passage of the labelled DNA as well as the unincorporated nucleotides were monitored by placing the output tube from the sephadex column onto a radiation monitor. The first peak of radiation, which contained the nick-translated DNA, was collected and stored at -20°C until needed.

An alternative method for preparing nick-translated DNA has involved the use of the "Nick-translation Kit" purchased from Bresa. The kit includes the non-radioactive nucleotides along with an enzyme mix containing DNA'ase and DNA polymerase I. Nick-translation is accomplished by adding 3  $\mu$ l of an aqueous solution of [ $\alpha$ -<sup>32</sup>P] dCTP (2'-deoxycytidine 5'-triphosphate, Amersham), 4  $\mu$ l of the nucleotide mix and 5  $\mu$ l of the enzyme mix to 8  $\mu$ l of the appropriate DNA template in H<sub>2</sub>O and incubating this reaction mixture at 14°C for 90 min. The processing of the nick-translated DNA is identical to that previously described except for the following modification. The nick-translated and ligated sample was not passed over a sephadex column, but rather, was subjected to two ethanol precipitations. Two hundred microlitres of TE was added to the pellet, the DNA was resuspended and then stored at -20°C.

## 2.5 Synthesis of $^3\text{H}$ c-RNA for in situ Hybridization Probes

The protocol used in these experiments involved a modification of the procedure reported by Pardue et al. (1970).

In an acid washed tube, 100  $\mu\text{l}$  each of  $^3\text{H}$  CTP ([5- $^3\text{H}$ ] cytidine 5'-triphosphate, 25 Ci/<sup>mmol</sup>, Amersham),  $^3\text{H}$  UTP ([5,6- $^3\text{H}$ ] uridine 5'-triphosphate, 40 Ci/<sup>mmol</sup>, Amersham) and  $^3\text{H}$  ATP ([2,8- $^3\text{H}$ ] adenosine 5'-triphosphate, 29 Ci/<sup>mmol</sup>, Amersham) along with 10mM GTP (guanosine 5'-triphosphate, P-L Biochem) were mixed and then evaporated using a rotary vacuum. These nucleotides were resuspended in 20  $\mu\text{l}$  of  $\text{H}_2\text{O}$  and transferred to an eppendorf tube containing 1  $\mu\text{g}$  of the selected DNA in 15  $\mu\text{l}$  of  $\text{H}_2\text{O}$ . To this tube were added 10  $\mu\text{l}$  of 5 X RNA polymerase buffer (40mM Tris HCl, pH 7.9, 10mM  $\text{MgCl}_2$ , 0.1mM EDTA, 150mM KCl, 500  $\mu\text{g}/\text{ml}$  bovine serum albumin), 0.5  $\mu\text{l}$  of 1M DTT and 5  $\mu\text{l}$  RNA polymerase (0.9 units/ $\mu\text{l}$ , Boehringer). The reaction mixture was incubated at  $37^\circ\text{C}$  for 30 min. at which time the percentage of incorporated  $^3\text{H}$  nucleotides was determined as described in section 2.4. Two hundred and fifty  $\mu\text{l}$  of 0.05M Tris, 10  $\mu\text{l}$  of DNA'ase (0.01 mg/ml, Boehringer), 5  $\mu\text{l}$  of yeast RNA (10 mg/ml) and 5  $\mu\text{l}$  of <sup>0.05M</sup>  $\text{CaCl}_2$  were added to the reaction solution and the mixture was allowed to incubate a further 30 min. at room temperature. The c-RNA reaction solution was subsequently extracted by adding 25  $\mu\text{l}$  of 0.25M EDTA, 1  $\mu\text{l}$  of 5% SDS, 100  $\mu\text{l}$  of phenol and 100  $\mu\text{l}$  of chloroform. The phenol/chloroform phase was washed with 100  $\mu\text{l}$  of 0.05M Tris and the pooled aqueous phases were loaded onto a sephadex 200 column. Twenty-four X 14 drop fractions were collected utilizing a Gilson Micro Fractionator. One  $\mu\text{l}$  aliquots were taken from each of the fractions and spotted onto

glass fibre filters. The filters were then assayed for the amount of  $^3\text{H}$  present and the first peak of radiation was collected for use in in situ hybridization experiments.

## 2.6 Restriction Endonuclease Experiments

Numerous analyses involving the use of restriction endonucleases were carried out during the course of this study (see Table 2-1 for enzymes used). In all reactions the following percentages of the various constituents were held constant: selected DNA sample, 66% of the final volume; appropriate 10 X buffer (Table 2-1), 10% of the final volume; 2-mercaptoethanol (1% solution), 6.6% of the final volume; appropriate enzyme (used at 2 units/ $\mu\text{g}$  of DNA unless otherwise noted) plus an amount of  $\text{H}_2\text{O}$  to equal 17.4% of the final volume. This mixture was incubated at the suitable temperature (Table 2-1) for 1 hr. The enzyme reaction was terminated by adding 5% SDS to a final concentration of 0.1% and then incubating the sample at  $65^\circ\text{C}$  for 3 min. Loading dye (50% glycerol, 0.12M EDTA pH 8.0, 0.5% SDS, 0.25% Bromophenol Blue) was added to each sample equivalent to 20% of the final volume and the sample was stored on ice until required.

## 2.7 Gel Electrophoresis

Depending upon the length of the DNA restriction fragments under investigation, either an agarose or acrylamide matrix was utilized. In addition, various percentages of agarose and acrylamide were employed in order to resolve different sizes of



Table 2-1. Buffers, incubation temperatures and suppliers for  
the restriction endonucleases used in this study.

<u>ENZYME*</u>	<u>BUFFER**</u>	<u>INCUBATION TEMP (°C)</u>	<u>SUPPLIER</u>
Taq I	3	65	R. Appels
Bam HI	4	37	Biolabs
Eco RI	1	37	Biolabs
Hind III	2	37	Biolabs
Pst I	4	37	Biolabs
Pvu I	4	37	Biolabs
Acc I	3	37	Biolabs
Ava II	3	37	Biolabs
Xma III	3	37	Biolabs
Rsa I	3	37	Biolabs
Hae III	3	37	Biolabs
Hha I	5	37	Biolabs
Sau 3A1	3	37	Biolabs
Cla I	3	37	Biolabs

\* See Roberts (1979) for a list of sources and references

\*\* 1=100mM Tris-HCl (pH 7.5), 50mM NaCl, 5mM MgCl<sub>2</sub>, 100 µg/ml bovine serum albumin;

\*\* 2=50mM Tris-HCl (pH 8.0), 50 mM NaCl, 10mM MgCl<sub>2</sub>, 100 µg/ml bovine serum albumin;

\*\* 3=6mM Tris-HCl (pH 7.4), 50mM NaCl, 6mM MgCl<sub>2</sub>, 6mM 2-mercaptoethanol, 100 µg/ml bovine serum albumin;

\*\* 4=6mM Tris-HCl (pH 7.9), 150 mM NaCl, 6mM MgCl<sub>2</sub>, 100 µg/ml bovine serum albumin;

\*\* 5=50mM Tris-HCl (pH 8.0), 5mM MgCl<sub>2</sub>, 0.5mM DTT, 100 µg/ml bovine serum albumin.

fragments. Agarose gels consisted of the suitable amount of agarose (Sigma, Type II: Medium EEO), 7.5 ml of 20 X electrophoresis buffer (0.8M Tris, 0.4M sodium acetate, 0.04M EDTA, pH 8.4), 142.5 ml of H<sub>2</sub>O and 0.15 ml of ethidium bromide (0.5 mg/ml, Sigma). Subsequent to dissolving the agarose with boiling, the solution was used for either horizontal slab gels or disc gels. The disc gels were employed in analyzing an aliquot of the samples in order to verify that complete digestion had occurred. These gels were electrophoresed at 10mA per gel. The running buffer for both the disc and slab gel electrophoresis consisted of 100 ml of the 20 X electrophoresis buffer, 2 ml of ethidium bromide and 1898 ml of H<sub>2</sub>O. The horizontal gels were electrophoresed at 100mA until the bromophenol blue had migrated approximately 75% of the gel length.

A vertical slab acrylamide system was also employed for higher resolution of restriction fragments. The percentage of acrylamide used in any one gel ranged from 8-12%. An 8% gel consisted of 27 ml of a 30% acrylamide solution (142.5 gm of acrylamide, Sigma, 7.5 gm of N,N'-methylene-bis-acrylamide-Sigma, in 500 ml H<sub>2</sub>O), 10 ml of 10 X TBE buffer (108 gm Tris, 9.3 gm EDTA, 55 gm Boric acid, up to 1 l with H<sub>2</sub>O, pH 8.3), 3.3 ml of a 1.6% solution of ammonium persulphate (BDH), 20 ml of a 50% glycerol solution plus H<sub>2</sub>O to bring the final volume up to 100 ml. This solution was filtered, by suction, through two no. 1 qualitative paper filters (Whatman) and was then left under vacuum for approximately 2 min to remove any air bubbles. Forty  $\mu$ l of N,N,N',N'-tetramethyl-ethylenediamine (TEMED, Sigma) was added to the acrylamide solution and the mixture was



then poured between two 390 mm X 200 mm glass plates with 0.3 mm spacers. Samples to be run on these gels were treated in the same manner as for agarose gels. The running buffer for these gels consisted of 1 X TBE (1/10 dilution of 10 X TBE).

Following electrophoresis, it was necessary to stain the acrylamide gel in a 0.005 mg/ml solution of ethidium bromide. Visualization of the electrophoresed samples were then possible (in both the acrylamide and agarose systems) by placement of the gels onto a light box equipped with ultraviolet lamps (15 watt). The DNA bands were then photographed with a mounted Polaroid Land Camera (MP4) using Polaroid Coaterless Land Pack film (Type 667 or 665).

## 2.8 Transfer of DNA from Agarose Gels to a Nylon Membrane

In numerous experiments, DNA fragments electrophoresed in an agarose matrix were transferred to Gene Screen (New England Nuclear) using a modification of Southern's (1975b) technique. Gels containing the DNA fragments were trimmed so that an approximately 1 cm border remained on each side. In addition, the cathodal portion of the gel was removed as well as the portion from 3 cm anodal to the bromophenol blue dye front. The remaining gel was immersed in a 1.5M NaCl, 0.5M NaOH denaturing solution and was stirred gently for 30 min. The denaturant was removed, the gel rinsed twice with approximately 200 ml of distilled H<sub>2</sub>O, neutralizing buffer (3M NaCl, 0.5M Tris-HCl, pH 7.0) was added and the solution stirred gently for at least 45 min. Prior to the denaturing step, a piece of Gene Screen (approximately 2mm greater in width and length than the gel

piece) was placed into  $H_2O$  and stirred for the entirety of the denaturing and neutralizing reactions. Four pieces of no. 3 chromatography paper (Whatman), which measured 1 cm greater in width and length than the gel piece, were prepared. Two of these pieces were laid onto plastic food wrap and subsequently soaked with 20 X SSC. The denatured and neutralized gel was inverted and placed onto the chromatography paper, ensuring that no air bubbles were trapped between the gel and the filter paper. Excess 20 X SSC was blotted away from the gel and paper. Parafilm strips were then placed along each of the four sides of the gel in order to cover the chromatography paper previously left exposed by the gel. The Gene Screen was applied such that one edge was aligned with the base of the gel (located at the loading wells) and one side aligned with one of the sides of the gel. One of the two remaining pieces of chromatography paper was soaked with water and placed onto the Gene Screen. The remaining piece of dry paper was laid onto the wet layer, followed by a 3 cm thickness of paper tissues. From one to four glass plates were placed on top of the tissues and then an approximately 1 Kg weight was positioned on the glass plates. The transfer was allowed to proceed for at least 16 hr at which time the layers were removed and the Gene Screen was separated from the gel and baked under vacuum for 2 hr at 80°C.

## 2.9 Hybridization of $^{32}P$ -Nick-Translated DNA to DNA Fragments Immobilized on Gene Screen

Nick-translated DNA (See Section 2.4) was added to a solution consisting of 0.5 ml of 5% SDS, 3.75 ml of 20 X SSC, 15

$\mu$ g of an appropriate competitor DNA (partially digested with the restriction endonuclease Hae III), 1.25 ml of 100 X Denhardt's solution (2% bovine serum albumin, Sigma; 2% Ficoll 400, Pharmacia; 2% Polyvinyl-pyrrolidone; Denhardt, 1966) and 12.5 ml of formamide (Fluka). The final volume was taken up to 25 ml with TE. Prior to hybridization, the Gene Screen containing the immobilized DNA (see Section 2.8) was wetted in hybridization solution, without any radioactive probe, for 1 hr with gentle stirring. This solution was then poured off. The hybridization solution containing the nick-translated DNA was incubated at 95°C for 5 min and then poured onto the wetted Gene Screen. The hybridization was routinely carried out at 37°C over a time period of from 4 to 16 hr. The filter was removed and washed at room temperature in 2 X SSC/0.1% SDS (solution stored at 65°C) for up to 4 hr with changes of the wash solution occurring at 30 min intervals. This filter was covered in plastic food wrap, covered with pre-flashed X-ray film, placed in an autoradiography cassette with an intensifying screen (Laskey, 1980) and stored at -70°C until developed.

Filters which had undergone hybridization were re-used by simply separating the nick-translated probe from the membrane bound DNA in the following manner. Two hundred and fifty ml of a solution containing 125 ml of formamide and 125 ml of TE was heated to 85-90°C. One half of this solution was placed in a container, the filter was added and the solution was stirred at room temperature for 10 min. The remainder of the formamide/TE solution was kept at 85-90°C and was added to the



container following the removal of the first 125 ml portion. The filter and solution were again allowed to stir for 10 min. The filter was then transferred to the 2 X SSC/0.1% SDS solution (see above) and washed for 15 min. Prior to hybridization, the filter was once again placed in hybridization solution, which lacked the radioactive probe, for 5-10 min and then treated as discussed above.

#### 2.10 Electroelution of DNA from Agarose Gels

In order to recover specific DNA fragments for further analysis, DNA was digested with endonucleases by the above method (see Section 2.6) and run on agarose gels, also described above (see Section 2.7). The area of gel containing the appropriate fragment was excised and placed in a collodion bag (Sartorius). Fresh electrophoresis running buffer, without ethidium bromide, was then added to the bag so that when the bag was placed in a horizontal position the gel piece was covered with the liquid. The collodion bag was placed in a 4 X 1.5 cm slot which had been cut in a 2% agarose gel. A layer of running buffer was added to the slot and then electrophoresis was carried out as described previously. The time necessary for the electroelution of a particular fragment was dependent on the DNA fragment size. Following electroelution, the liquid was pipetted from the collodion bag and the bag and gel were rinsed with another aliquot of the running buffer. The two aliquots were pooled and an equal volume of isoamyl alcohol (saturated with H<sub>2</sub>O) was added. This mixture was vortexed and the two layers were allowed to separate. The alcohol layer was discarded and the remaining

aqueous layer was transferred to a polyallomer centrifuge tube (Beckman). The sample was ethanol precipitated and centrifuged at 27,000 rpm for 1 hr in a Beckman L8-80 Ultra-centrifuge. The liquid was decanted, the DNA pellet was dried, resuspended in TE buffer and stored at -20°C until needed.

## 2.11 Cloning Caledia Highly Repetitive DNA Using the Plasmid

### Vector pBR322

Linearized pBR322 was prepared by digesting the plasmid with the appropriate restriction endonuclease. The enzyme was chosen so that the two ends of the restricted plasmid were complementary to the ends of the DNA fragment to be cloned. Following digestion, a 0.5 µg aliquot of the DNA was electrophoresed on an agarose disc gel to confirm that complete digestion had taken place. One hundred µl of phenol and chloroform, along with 50 µl of TE were added and the sample was processed as described in Section 2.4. The plasmid DNA was resuspended in H<sub>2</sub>O such that the final concentration was 0.5 µg/µl. Formation of recombinant plasmids was accomplished by ligating the selected C. captiva DNA fragment with the linearized pBR322. The ligation was carried out as described in Section 2.4 using equivalent amounts of the vector and insert DNA. Just prior to the transformation experiments, 1 µl of the ligation mixture was added to 30 µl of CaCl<sub>2</sub> (0.05M), mixed well and placed on ice.

Two strains of Escherichia coli (ECR291 and RR1) were utilized for transformation via CaCl<sub>2</sub> treatment (Lederberg and Cohen, 1974). Cells from both of these strains were processed as described below. An overnight culture of the bacterial cells was

grown in 50 ml of LB medium (10 gm tryptone, Difco; 5 gm yeast extract, Difco; 5 gm NaCl; 2 gm glucose; 1 ml 1M  $\text{MgCl}_2$ ; taken up to 1 l with  $\text{H}_2\text{O}$ ) at 37°C with shaking. One ml of the overnight culture was placed in 100 ml of fresh LB medium and grown until the cells had reached an OD600 of approximately 0.7. Two 10 ml and one 5 ml aliquot of cells were transferred to sterile, plastic centrifuge tubes and centrifuged at 2000 rpm for 5 min. The medium was decanted and the remaining cells were resuspended in cold (approximately 5°C)  $\text{MgCl}_2$  (0.01M). The cells were then combined into two of the centrifuge tubes. A further amount of  $\text{MgCl}_2$  was added to the tubes to take the final volume to 10 ml per tube. The cells were again centrifuged at 2000 rpm for 5 min, the pellets were resuspended in 6.25 ml of cold  $\text{CaCl}_2$  (0.05M) and placed on ice for 20 min. Subsequently, the cells were centrifuged again and the remaining pellets were resuspended in 6.25 ml of  $\text{CaCl}_2$ . The cell suspensions were pooled and stored at 5°C until use. One hundred ml of this cell suspension was added to the ligation mixture in  $\text{CaCl}_2$  (see above) and this sample was placed on ice for 1 hr. This solution was incubated at 42°C for 2 min and then returned to an ice bath. LB medium (0.7 ml) was added to the sample and this mixture was incubated for 1 hr at 37°C. The samples were plated out onto fresh ML agar plates (10 gm glucose, 5 gm NaCl, 5 gm yeast extract, 10 gm tryptone, 20 gm agar; taken up to 1 l with  $\text{H}_2\text{O}$ ) which contained the appropriate antibiotic. One half of the transformed cells were placed on an ML plate and the mixture was spread across the plate with the aid of a glass rod. Once an even distribution of the mixture had been achieved, the plate was



placed upright with the lid slightly ajar in a 37°C incubator. This plate was closed after the surface of the agar was dry. The remainder of the transformed cells were processed in the same fashion using a second ML plate. The dry plates were inverted and left at 37°C until the bacterial colonies could be easily seen.

The colonies were initially screened for the presence of inserted DNA by taking advantage of the presence of antibiotic resistance genes present in the pBR322 plasmid. Insertion of foreign DNA inactivates one of these genes by its insertion into either the gene or its promoter region. Therefore, the colonies were transferred, with sterile toothpicks, from the original plate to a plate containing an antibiotic to which the cells were resistant and then to a second plate that possessed an antibiotic that would kill the cells. The selected colonies were then grown in liquid cultures (see below) and the plasmid DNA recovered was subjected to gel electrophoresis, transfer to Gene Screen and hybridization to an appropriate <sup>32</sup>P-nick-translated DNA (see Section 2.8 and 2.9). Alternatively, the colonies were transferred to a plate which contained an antibiotic to which the cells were resistant. The colonies were placed according to a grid pattern to facilitate identification. After allowing the colonies to grow for approximately 12 hr, a membrane filter (Schleicher and Schull) was placed on top of the colonies and then lifted off to remove samples from each colony. The filter was placed, with the colonies on top, onto a plate containing the same antibiotic and the bacteria were allowed to

grow once more. The filters were transferred to an ML plate which contained chloramphenicol (Sigma) and remained there for another 12-16 hr. The chloramphenicol inhibited cell division, but allowed the replication of the plasmid DNA. Following this treatment, the filters were removed from the plate and placed onto no. 4 filter paper (Whatman) to which 5 ml of denaturant solution had been added. The filters were left for 10 min, laid onto paper towelling to remove excess denaturant and then were placed onto a second piece of filter paper. Five ml of neutralizer solution was added to this filter paper. The filters were left on this paper until they were neutral. The filters were dried on paper towelling and then processed as described for "Southern" filters (see Sections 2.8 and 2.9) in order to determine which colonies had plasmids containing the highly repeated DNA inserts.

Isolations of plasmid DNA were carried out in various amounts of media ranging from 50-1000 ml. The experimental methods described below were those used for 1 l preparations. One method of preparing plasmid DNA involved the use of M9 medium. This medium consisted of 100 ml of 10 X M9 (60 gm  $\text{Na}_2\text{HPO}_4$ , 30 gm  $\text{KH}_2\text{PO}_4$ , 5 gm  $\text{NaCl}$ , 10 gm  $\text{NH}_4\text{Cl}$ ; up to 1 l with  $\text{H}_2\text{O}$ ), 20 ml of 20% glucose, 10 ml of 0.1M  $\text{MgSO}_4$ , 10 ml 0.01M  $\text{CaCl}_2$  and 500 ml of 1% casamino acids/thiamine HCl (5.0 gm casamino acids, Difco; 0.002 gm thiamine HCl, Sigma; up to 500 ml with  $\text{H}_2\text{O}$ ). The appropriate antibiotic was added to this medium after autoclaving. Ten ml of this solution was placed in a sterile, capped test tube and part of the selected bacterial colony was transferred to the tube using a flamed metal loop.

This culture was incubated for approximately 16 hr at 37°C with shaking. The entire culture was transferred to 1 l of M9 medium and incubated at 37°C with shaking until the cells reached OD600 = 0.1. Uridine (0.05 gm; Sigma) was added and the cells were allowed to grow to an OD600 = 0.6, at which time chloramphenicol (0.1 gm) was added. The culture was then grown (37°C) for up to 20 hr. In order to harvest the bacterial cells, the media were centrifuged for 20 min at 6000 rpm in 500 ml plastic centrifuge bottles (Nalgene). The supernatant was discarded and the two pellets were resuspended in 200 ml of cold TE. This suspension was centrifuged again at 6000 rpm for 20 min. The supernatant was discarded, the pellets were resuspended in 25 ml of cold sucrose solution (25% in 0.05M Tris-HCl, pH 8.0), pooled, placed on ice and shaken vigorously for 10 min. The following solutions were successively added at 10 min intervals: 5 ml of cold lysozyme (10 mg/ml; Sigma); 5 ml of EDTA (0.5M, pH 8.5); and 40 ml of a triton solution (1 ml 10% triton, Univar; 12.5 ml 0.5M EDTA pH 8.5; 5 ml 1M Tris-HCl pH 8.0; 80 ml H<sub>2</sub>O). The cell solution was stirred for an additional 10 min, transferred to plastic centrifuge tubes (Beckman) and centrifuged at 27,000 rpm for 30 min. The supernatant was poured away from the resulting pellet into a 100 ml graduated cylinder. The volume was noted and 0.95 gm of CsCl (Ajax Chem) per ml of solution along with 1 ml of EtBr (10 mg/ml) per 10 ml of solution was added. The CsCl was dissolved by gentle inversion of the cylinder. This mixture was decanted into quick seal polyallomer centrifuge tubes (Beckman) and centrifuged for at least 20 hr at 44,000 rpm.



Following centrifugation, the DNA bands were visualized with UV irradiation. The lower, or plasmid, band was removed from the tube by suction with a needle and hypodermic syringe. The plasmid sample was extracted six times with equal amounts of isopropanol saturated with NaCl. Following these extraction steps, the DNA solution was dialyzed in 2 l of TE with three changes at 1 hr intervals. The DNA was ethanol precipitated and subsequently resuspended in TE buffer to a final concentration of approximately 1 µg/µl. Occasionally, the plasmid preparations were also treated with RNA'ase prior to the ethanol precipitation step. This was accomplished by adding 0.2 ml of 12 X SSC per 1 ml of plasmid solution and 10 µl of RNA'ase A (10 mg/ml) per 1 ml of solution, followed by a 37°C incubation for 30 min. The solution was extracted with phenol and chloroform and then ethanol precipitated. The plasmid DNA, in TE buffer, was stored at -20°C when not being used.

A more recent method of isolating plasmid DNA has involved the use of LB medium in place of M9 medium. The only modification to the "M9" method was that following the addition of the 10 ml culture to the LB medium, the cells were grown to an OD600 = 1.0 and the chloramphenicol was added. The culture was then treated as described above.

## 2.12 Cloning Caledia Highly Repetitive DNA into the Vector Phage λ1059

The preparation of DNA for cloning into phage λ, the isolation of λ 1059 arms and the cloning of the recombinant phage have been described by Appels and Moran (1984).

The screening of the recombinant phage for clones containing Caledia highly repeated DNA was accomplished by utilizing a modified form of the filter lift-off procedure described in Section 2.11. The filters were laid onto the plates containing phage and were subsequently removed, denatured, neutralized, fixed, prehybridized and hybridized as described previously. Plaques containing highly repeated sequences were removed from the plate with a pasteur pipette and transferred to a tube with 1 ml of <sup>phage</sup> storage buffer (PSB; 5 ml 1M Tris pH 7.4, 2.9 gm NaCl, 0.05% gelatin in 500 ml of H<sub>2</sub>O) and 10 µl of chloroform. This was vortexed and allowed to sit on ice for 15 min. Fifty µl of this solution was added to a tube containing 50 µl of PSB and 1 µl of E. coli C-600 cells. The C-600 cells were prepared in the following manner. A single colony of C-600 cells was grown to an OD<sub>600</sub> in λ LB medium (10 gm tryptone, 5 gm yeast extract, 5 gm NaCl, 10 ml 1M Tris pH 7.4, 2 gm maltose). The cells were centrifuged for 10 min at 3000 rpm, washed with cold 0.01M MgCl<sub>2</sub> and pooled in a single tube. The cells were centrifuged once again and subsequently resuspended in 0.4 ml of PSB.

The tube containing the PSB, C-600 cells and the phage solution was placed at 37°C for 15 min. This entire sample was then pipetted into 70 ml of NZ-amine medium (10 gm NZ-amine, Difco; 5 gm NaCl; 5 gm yeast extract; 1 gm casamino acids; 2 gm MgCl<sub>2</sub>; up to 1 l with H<sub>2</sub>O) which had been prewarmed to 37°C. The culture was incubated at 37°C, with shaking, until the bacterial cells lysed from the phage infection. At that time, 0.2 ml of chloroform was added to the culture flask and the mixture was

innoculated for a further 30 min. NaCl (2-3 gm) was added to the culture and dissolved. The culture was placed on ice for at least 1 hr and was then centrifuged at 8000 rpm for 10 min. The supernatant was collected and subsequently centrifuged for 3 hr at 35,000 rpm. The pellet was resuspended in PSB and the volume was increased to 5.7 ml with PSB. Four gm of CsCl were added to the solution and then dissolved. The gradient was centrifuged for 20 hr at 35,000 rpm. The phage were collected from the gradient, diluted with PSB and centrifuged for an additional 3 hr at 35,000 rpm. The pellet was resuspended in 300  $\mu$ l of TE and then 300  $\mu$ l of extraction mix (200  $\mu$ l chloroform; 200  $\mu$ l of phenol; 100  $\mu$ l  $\lambda$  lytic extraction buffer, 30mM EDTA, 100mM Tris pH 7.5, 10 mM NaCl, 1% sarkosyl) were added. This mixture was vortexed vigorously and the aqueous phases removed and pooled. The DNA was ethanol precipitated, resuspended in 100  $\mu$ l of TE buffer and stored at -20°C.

The  $\lambda$  clones discussed in this thesis were originally isolated by R. Appels and F. Clay.

### 2.13 Cloning of Caledia Highly Repeated DNA into the Vector M13

Linear M13 mp8, mp9 and mp10 phage (Messing, 1981; Messing and Vierira, 1982) were prepared in the same manner as the pBR322 vectors (see Section 2.11). To increase the efficiency of producing recombinant phage, the linear DNA was treated with the enzyme alkaline phosphatase to remove the 5'-terminal phosphates from the linearized phage (Ullrich et al., 1977) and thus prevent the religation of the phage DNA without an insert. The digested DNA was ethanol precipitated and then resuspended in 40  $\mu$ l of



H<sub>2</sub>O. Alkaline phosphatase (Boehringer; 0.025 units per 1 $\mu$ g of vector DNA), along with 10 X alkaline phosphatase buffer (0.1M glycine, 0.001M MgCl<sub>2</sub>, 0.0001M ZnCl<sub>2</sub> pH 10.5) were added to the DNA to make a 1 X final solution and this solution was incubated at 37°C for 30 min. The reaction was stopped by adding 5% SDS (to make a 0.1% final concentration) and the sample subsequently incubated for 45 min at 68°C. The solution was allowed to cool to room temperature and was extracted with phenol and chloroform. The aqueous phase was ethanol precipitated and resuspended in H<sub>2</sub>O to a final concentration of 0.5  $\mu$ g/ $\mu$ l.

Following the ligation (see Section 2.11) of the vector and Caledia DNA, 1  $\mu$ l of the dilution mixture was combined with 25  $\mu$ l of CaCl<sub>2</sub> and 0.3 ml of E. coli JM 101 competent cells. The JM 101 competent cells were prepared as described in Section 2.11 except that they were grown in 2YT medium (16 gm tryptone, 10 gm yeast extract, 5 gm NaCl, up to 1 l with H<sub>2</sub>O). The ligation solution/CaCl<sub>2</sub>/JM101 mixture was placed on ice for 40 min, incubated at 42°C for 2 min and then returned to the ice bath. Ten  $\mu$ l of Isopropyl  $\beta$ -D-Thio-galactopyranoside (IPTG, 0.1M; Sigma), 50  $\mu$ l of 5-Bromo-4-Chloro-3-Indolyl- $\beta$ -D-Galactopyranoside (X-gal, 2% in dimethyl formamide; Sigma), 0.2 ml of exponentially growing JM 101 cells and 3 ml of H-top agar (0.65% agar in H medium, see below) were added to the sample. This mixture was poured onto an H plate (10 gm tryptone, 8 gm NaCl, 12 gm agar), allowed to harden and was then placed at 37°C until plaques were visible (approximately 16 hr). The screening of the recombinant plaques were carried out using the filter

lift-off technique described in Section 2.12. Single-strand DNA, for use in DNA sequencing protocols (see below), was isolated from clones which were demonstrated to have Caledia highly repeated DNA inserts. The selected plaques were removed from the H plate with a pasteur pipette and transferred to a sterile, capped test tube which contained 2 ml of exponentially growing JM 101 cells. The cultures were grown from 6 - 16 hr at 37°C with shaking. A 1.5 ml aliquot was transferred from the test tube to an eppendorf centrifuge tube and was centrifuged for 5 min. The supernatant (1.2 ml aliquot) was placed in a second eppendorf tube along with 2 µl of RNA'ase A (1 mg/ml) and 1 µl of DNA'ase (0.2 mg/ml). This mixture was allowed to incubate at room temperature for 30 min. Two hundred µl of a polyethylene glycol 6000 (PEG, 20%; BDH Chem) and NaCl (2.5M) solution was added and the sample was placed on ice for 30 min. The solution was centrifuged for 5 min and the supernatant was removed by aspiration. The pellet was resuspended in 100 µl of TE and then 50 µl of phenol were added. Following vortexing, the sample was allowed to stand at room temperature for 10 min and was once again subjected to vortexing. The layers were separated by a 5 min centrifugation and the aqueous layer removed. This layer was ethanol precipitated and taken up in 50 µl of TE. In order to verify that the DNA isolation had a Caledia highly repeated fragment, an aliquot of the sample was electrophoresed on an agarose gel, transferred to Gene Screen and hybridized to the appropriate probe.

For further analysis it was sometimes necessary to have an isolated M13 clone in the form of double-stranded DNA.

Therefore, the double-stranded replicative form was isolated in the following manner. The single-stranded phage was used to transform JM 101 cells. Following this, a single plaque from the plate was added to 2 ml of 2YT medium in a sterile, capped test tube and grown for 10 hr at 37°C. Ten ml of 2YT was inoculated with a single colony of JM 101 and grown for 12.5 hr at 37°C. The JM 101 culture was then added to 1 l of 2YT and allowed to incubate for a further 2.5 hr at which time the 2 ml phage inoculum was added. This combined culture was grown for 4-5 hr and was then processed in the same manner as the pBR322 preparations (see Section 2.11).

## 2.14 Melting Point ( $T_m$ ) Determinations and Saturation

### Hybridization Experiments

Melting point determinations were carried out as described by Appels and Dvorak (1982). DNA (10  $\mu$ g) samples were denatured in 0.2 mls of denaturant (37°C, 10 min) and cooled on ice. Cold neutralizer solution was added and the mixture immediately loaded onto 25 mm discs of nitro-cellulose (Sartorius). The DNA was fixed onto the filters at 80°C/2hr. The filters were prehybridized for 1 hr in hybridization mix without any radioactive probe and then hybridized at 37°C for 10-12hr followed by a wash in 2 X SSC/0.1% SDS at 65°C. For the  $T_m$  determinations of the hybridized probe the filters were serially transferred to 0.75 ml aliquots of 3 X SSC/50% formamide at the required temperature. The filters were maintained at the required temperature for 5-8 min and then transferred to the next



\*Three aliquots (25 $\mu$ g, 5 $\mu$ g and 1 $\mu$ g) of DNA were used to determine the copy number of specific highly repeated sequences in C. captiva populations. These aliquots were prepared in the same manner as those used for the melting point determinations, except that they were loaded onto 45mm nitrocellulose filters. Subsequent to the prehybridization step, 15 7mm discs were cut from each of the 45mm filters. Three of the discs were placed into five serial dilutions (1 : 0.5 : 0.25 : 0.125 : 0.0625) of a hybridization solution containing a specific nick translated sequence. Following hybridization, the individual discs were placed in scintillant and the amount of radioactivity was determined.

The inverse of the serial dilution (X - axis) was then plotted against the inverse of the radioactivity present on the filter. The Y-intercept was determined and the inverse of this value was divided by the specific activity of the nick translated sequence. The resulting value (representing the amount of the sequence present in the sample) was divided by the total DNA present on the 7mm disc; this yielded the percentage of the genome represented by the sequences. This value was multiplied by the haploid amount of base pairs in the C. captiva genome (Rees et al., 1978). The subsequent value was then divided by the length, in base pairs, of the sequence being assayed in order to define its copy number.

vial. Up to twenty melting point determinations were carried out at the same time, with the  $^{32}\text{P}$ - probe hybridized to a specific DNA sample used as an internal standard for each experiment. To determine the radioactivity released at the various temperatures 0.35 ml of  $\text{H}_2\text{O}$  and 9 ml of Triton-X based scintillation fluid were added to the vials.

Saturation hybridization experiments to determine the copy number of the repetitive sequence followed the procedure of Bishop et al. (1969). \*

## 2.15 In situ Hybridization of $^3\text{H}$ -c-RNA from Caledia Highly Repeated DNA to Prepared Chromosomes

Air dried preparations of Caledia chromosomes were prepared as described by Webb (1976). These preparations were then utilized for in situ hybridization experiments as described by Appels et al. (1978) with some modifications. The microscope slides on which the chromosomes had been prepared were taken through a series of ethanol washes (50%-70%-100%). The slides were air dried and then taken through 70% ethanol and 2 X SSC and finally placed in 2 X SSC with 10  $\mu\text{g}/\text{ml}$  RNA'ase A for 30 min at room temperature. The slides were removed and passed through 2 X SSC, two  $\text{H}_2\text{O}$  washes and then placed in 0.2 N HCl at  $37^\circ\text{C}$  for 10 min. Subsequently, the preparations were washed in two lots of sterile  $\text{H}_2\text{O}$ , two of 70% ethanol, one 100% ethanol wash and placed in 100% ethanol for 5 min. The slides were removed and allowed to dry. The  $^3\text{H}$ -c-RNA (see Section 2.5) was resuspended in a 3 X SSC and 50% formamide solution such that there were  $10^4$ - $10^5$  cpm per  $\mu\text{l}$  of solution. Seven  $\mu\text{l}$  of this solution

was added to each slide. A coverslip was placed on top of the liquid and then was sealed with rubber cement. The cement was allowed to dry and then the slide was immersed in a 60°C H<sub>2</sub>O bath for 10 sec. The samples were allowed to hybridize at 37°C for approximately 20 hr. The glue and coverslip were removed and the slide was placed in a 3 X SSC and 50% formamide solution for 10 min at 37°C. The slide was washed through 2 X SSC and was then placed in an RNA'ase A solution for 30 min at room temperature. After this, the preparations were washed in 2 X SSC, 0.05% SDS or 2 X SSC, 0.1% SDS <sup>65°C cooling to</sup> at room temperature. The SSC, SDS solution was changed five times at 30 min intervals. The slides were rinsed through 2 X SSC, two lots of 70% ethanol, 100% ethanol and were then allowed to air dry. Nuclear research emulsion (Ilford, K2) was prepared by incubating equal amounts of the solid emulsion and H<sub>2</sub>O at approximately 45°C until the emulsion liquified. The H<sub>2</sub>O and emulsion were mixed and the slides were coated with the mixture. The emulsion was allowed to dry and then the slides were stored in the dark at 5°C. When the required amount of exposure time had transpired, the slides were developed (Kodak Dektol developer), fixed (Kodak fixative) and then stained with giemsa (Gurr) stain solution (1 ml NaPO<sub>4</sub> buffer, pH 6.8; 5 ml giemsa; taken up to 50 ml with H<sub>2</sub>O) for 10 min at room temperature. The slides were rinsed in H<sub>2</sub>O, dried and viewed under a light microscope. Photomicrographs were taken with Kodak technical pan film (2415).



## 2.16 DNA Sequencing

### 2.16.1 Dideoxy Chain Termination Method

A modified protocol of the DNA sequencing method described by Sanger et al. (1977) was utilized. The single-stranded M13 clones containing Caledia highly repeated DNA were used as templates for the sequencing reaction. Approximately<sup>e</sup> 1-2 µg of this DNA (5 µl) was mixed in an eppendorf tube with 1 µl of a 17 bp M13 primer (0.0025 µg/µl, Biolabs), 1.5 µl of a 10 X Klenow reaction buffer (100mM Tris, pH 8.5; 10 mM MgCl<sub>2</sub>) and 2.5 µl of H<sub>2</sub>O. This mixture was incubated at 55-60°C for 100 min. Adenosine, guanosine, cytosine and thymidine reaction mixtures were prepared as shown in Table 2-2. Four µl of these mixtures, along with 2 µl of the annealing mix, were placed in four separate tubes. This solution was incubated at 30°C for 15 min at which time 2 µl of the chase solution (equal volumes of 0.5mM dATP, dCTP, dGTP and dTTP) was added to each tube. The tubes were once again placed at 30°C for 15 min. At the end of the sequencing reaction, the samples were stored on ice for a maximum of 2-3 hr. Prior to loading the samples onto the sequencing gel, 4 µl of a formamide dye solution (0.1% xylene cyanol FF, 0.1% bromophenol blue, 10mM EDTA, 95% formamide) was added and the samples were incubated for 5 min at 95°C.

The glass gel plates used in these experiments were soaked for 1 hr in hot H<sub>2</sub>O containing KOH and detergent, rinsed and allowed to air dry in a fume hood. Absolute ethanol was applied to the plates and was then wiped off with tissues. This was repeated three times. The "gel" side of one plate was treated

Table 2-2. Reaction mixtures for the chain termination sequencing method. The deoxy solutions are those listed in Section 2.4.

ddATP = 2',3'-dideoxyadenosine 5'-triphosphate, 1 mM.

ddCTP = 2',3'-dideoxycytidine 5'-triphosphate, 0.3 mM.

ddGTP = 2',3'-dideoxyguanosine 5'-triphosphate, 0.1 mM.

ddTTP = 2',3'-dideoxythymidine 5'-triphosphate, 1.5 mM.

dCTP Mix = 5  $\mu$ l of 0.005 mM dCTP and 10  $\mu$ l of [ $\alpha$ -P<sup>32</sup>] (5  $\mu$ Ci/ $\mu$ l, Amersham).

Klenow activity = 5 units/ $\mu$ l (Boehringer).

Amounts shown below are in microlitres.

	<u>ADENOSINE</u>	<u>CYTIDINE</u>	<u>GUANOSINE</u>	<u>THYMIDINE</u>
d	8	8	8	8
dd	2	2	4	2
H <sub>2</sub> O	3	-	1	3
dCTP mix	3	6	3	3
Klenow	1	1	1	1



with a solution containing 50  $\mu$ l of methacryloxypropyltrimethoxysilane (Sigma), 10 ml of 95% ethanol and 300  $\mu$ l of 10% acetic acid. This solution was poured onto the plate and then wiped onto the entire surface until dry. The plate was washed three additional times with absolute ethanol and allowed to dry. The second plate was washed with ethanol and then approximately 10 ml of a 1% dimethyldichlorosilane (Sigma) solution was placed onto the plate. This solution was rubbed onto the entire surface until it was dry. This plate was washed with ethanol and allowed to dry. Plastic spacer strips were placed along the outside edges of one plate and the second plate was positioned on top of these. The plates were then taped together along the two sides and the bottom.

The sequencing gel consisted of 63 gm of urea (Univar), 40 ml of a 30% acrylamide solution (see Section 2.7), 15 ml of 10 X TBE (see Section 2.7) and 15 ml of  $H_2O$ . The mixture was heated at 45°C until the urea was dissolved and was then transferred to a 250 ml graduated cylinder. Ammonium persulphate (BDH Chem) was added (9.8 ml of a 1.6% solution) and the final volume was taken up to 150 ml with  $H_2O$ . The solution was aspirated through two layers of no. 1 filter paper (Whatman) and subsequently left under vacuum until all air bubbles had been removed. The gel solution was poured into a beaker containing 100  $\mu$ l of TEMED and was immediately poured into the mold prepared using the gel plates. A siliconized "comb" was pushed into the top of the gel to form the sample wells and the gel was left to polymerize for 1 hr. Following polymerization, the comb was removed and the wells were washed out with 1 X TBE. The gel plates were then placed

into an electrode chamber and the gel was pre-run at 2500 volts/30mA until the outside of the glass plates reached a temperature of 40°C. The wells were rinsed out with a syringe containing 1 X TBE to remove any urea and the reaction samples were loaded and electrophoresed. Subsequent to the electrophoresis, the tape was removed and the gel plates were split apart. The gel and glass plate were placed into a 10% acetic acid bath for 20 min. The plate and gel were removed from this bath and the gel was washed with a continuous flow of cold tap H<sub>2</sub>O for 10 min. The gel was dried onto the plate in a cabinet clothes dryer. When the gel surface was touch dry, X-ray film was placed onto the gel, the second gel plate was placed on top and the entire assemblage was wrapped with black plastic.

#### 2.16.2 Base-specific Chemical Cleavage Method

The sequencing protocol used for these experiments is that of Maxam and Gilbert (1980).

Preparation of the gel plates and the processing of the gel following electrophoresis was as described in Section 2.16.1 except for the following modifications. Both plates were treated with the 1% dimethyldichlorosilane solution. Following fixation, two pieces of filter paper were laid down such that they abutted against each other with one covering the top half and the other covering the bottom half of the gel. The gel was cut at the junction of these two pieces of filter paper and the gel pieces were lifted off in contact with the filter paper. The gel and paper was placed onto a Savant gel dryer and left until totally dry. Both pieces of the gel were covered with plastic wrap,

placed in an X-ray cassette, covered with X-ray film and placed at  $-70^{\circ}\text{C}$ .

#### 2.17 Isolation of Total RNA

Total RNA was extracted from adult grasshoppers. The grasshoppers were ground as described in Section 2.2. The resulting powder was placed in two sterile 30 ml Corex centrifuge tubes. Ten ml of a 1:1 mixture of NTES (0.1M NaCl, 0.01M Tris pH 7.5, 1mM EDTA, 0.5% SDS) and PCI (50 parts phenol, 48 parts chloroform, 2 parts iso-amyl alcohol) were added to each of the tubes and the solution was incubated for 10 min at  $65^{\circ}\text{C}$ . The samples were centrifuged at 8000 rpm for 15 min and the aqueous phases were transferred to another set of centrifuge tubes. An additional five ml of NTES was then added to the first set of tubes, the solution was mixed gently, the mixture was centrifuged and the supernatants were added to the previous aliquots. These solutions were ethanol precipitated and the pellets were resuspended on ice with mixing for 1 hr. Ten molar LiCl (1.25 ml) was added to each tube and the resulting mixture was left on ice for up to 48 hr. The solutions were centrifuged at 9000 rpm for 20 min, at which time the supernatant was discarded. The pellets were resuspended in 3.75 ml of  $\text{H}_2\text{O}$  and subsequently ethanol precipitated. The pellets were then resuspended in 0.5 ml of  $\text{H}_2\text{O}$  and the samples pooled.

#### 2.18 Electrophoresis of RNA Samples

RNA was electrophoresed on 1% agarose made up with 2 gm of agarose, 20 ml of 10 X running buffer (200 ml 1M 2-(N-morpholine)



ethane sulphonic acid; 40 ml 0.25M EDTA pH 7.5; 25 ml 2M sodium acetate; up to one litre with H<sub>2</sub>O; pH 7.0) and 1444.4 ml of sterile H<sub>2</sub>O. When the gel had cooled to 65°C, 35.6 ml of 36% formaldehyde was added. The gel plate and comb were wiped off with 0.5M NaOH, rinsed with H<sub>2</sub>O and the gel was poured. After the gel had polymerized, the comb was removed and the slots were immediately filled with 1 X running buffer. The RNA samples (10-20 µg) were placed in an eppendorf tube with an equal volume of sample buffer (100 µl 10 X running buffer, 50 µl formamide, 178 µl formaldehyde, 222 µl H<sub>2</sub>O). The mixture was incubated at 65°C for 5 min and then transferred to ice. A volume of loading dye (500 µl formamide, 500 µl glycerol, approximately 10 mg of bromophenol blue), equal to the sample buffer volume, was added and the entire sample was loaded onto the gel. A peristaltic pump was used to circulate the buffer from the anodal and cathodal electrode chambers.

#### 2.19 Transfer of RNA from an Agarose Gel to a Nitrocellulose Filter and Hybridization of Radioactive Probes to the Immobilized RNA

The transfer of the electrophoresed RNA onto a filter and the subsequent hybridization of radioactive probes to the RNA were carried out as described in Section 2.9 for DNA samples except for two modifications. First, the bottom filter paper, wetted with 20 X SSC, was kept in contact with the 20 X SSC solution throughout the transfer period. Secondly, a nitrocellulose membrane (BA-85, Schleicher and Schull) was used in place of Gene Screen.

## 2.20 Computer Assisted Analysis of the Repeated Sequences

Comparisons between repeats to determine the longest unbroken stretch of homology were made using the algorithm developed by Dromey (1979).

CHAPTER 3: The Heterochromatin of *Grasshopper* and *Cricket* *antenna* *Species* Complex. I. Sequence Evolution and Correlation in a Highly Repeated DNA Family.

This chapter has been accepted for publication in *Molecular Biology and Evolution*, and is in the format required for the journal. The authors are M.A. Arnold, S. Dromey, and J. D. D. Dromey.

Abstract.

The restriction enzyme *Xba*I digests 1-2% of the genomic DNA from the grasshopper *Cryptotympha* sp. in a family of sequences 162 bp in length (length of consensus sequence 162 bp). The sequence variation of this "Tag family" of repeat units was examined among several taxa within the *C. cryptotympha* species complex in order to assess the pattern of evolution within this highly repeated DNA. The Tag family repeats are located in C-banded heterochromatin on

**CHAPTER 3: The Heterochromatin of Grasshoppers from the Caledia captiva Species Complex. I. Sequence Evolution and Conservation in a Highly Repeated DNA Family**

This chapter has been accepted for publication in Molecular Biology and Evolution, and is in the format required for this journal. The authors are M.L. Arnold, R. Appels and D.D. Shaw.



**Abstract.**

The restriction enzyme Taq I digests 1-2% of the genomic DNA from the grasshopper Caledia captiva to a family of sequences 168 bp in length (length of consensus sequence). The sequence variation of this "Taq family" of repeat units was examined among several taxa within the C. captiva species complex in order to assay the pattern of evolution within this highly repeated DNA. The Taq family repeats are located in C-banded heterochromatin on at least one member of each homologous pair of chromosomes; the locations range from centromeric to telomeric. Thirty-nine cloned repeats isolated from two population 1A individuals along with 11 clones from seven populations taken from three taxa demonstrated sequence variation at 72 positions. Pairwise comparisons of the cloned repeats, both within an individual and between different subspecies, indicates that levels of intraspecific divergence, as measured by reproductive incompatibility, do not correlate with sequence divergence among the 168 bp repeats. A number of sequences within the repeat remain unchanged among all 50 clones; the longest of these sequences is 18 bp. The presence of the same 18 bp sequence in all clones examined departs significantly from random expectation ( $P < 0.01$ ). Two other cloned repeats from the sibling species of C. captiva have sequences that show 56% identity with this 18 bp conserved region. An analysis of the frequency of occurrence of a Rsa I recognition site within the 168 bp repeat in the entire Taq family agreed with that found in the cloned sequences. These data along with a partial sequence for the entire Taq family, obtained by sequencing uncloned repeats, suggest that the

consensus sequence from the cloned copies is representative of this highly repeated family and is not a biased sample resulting from the cloning procedure. The 18 bp conserved sequence is part of a 42 bp sequence that possesses dyad symmetry typical of protein binding sites, we speculate that this may be significant in the evolution of the Taq family of sequences.

### **Introduction.**

Numerous families of highly repeated DNA have been examined in plant and animal species (see Appels and Peacock, 1978; John and Miklos, 1979 for reviews). These studies have identified two major classes of highly repeated sequences, namely, those sequences that are interspersed within a genome and those that are tandemly repeated. The latter sequences are often referred to as satellite sequences because of the characteristic satellite band that they can form when centrifuged in certain buoyant density gradients. These sequences are tandemly repeated as many as  $10^6$  times within a genome and can be located on either a few or all of the chromosome pairs in a complement. The sequences are typically associated with heterochromatic regions of chromosomes. In addition, the sequence families that have been analyzed are often defined by the presence of a specific endonuclease recognition site located within each of the tandem repeats. The regular spacing of such sites within the tandemly arranged repeats suggests that homogeneity exists among the sequences within a given family. However, recent studies have revealed that extensive heterogeneity can be present within an

apparently homogeneous family of repeats (Lee and Singer, 1982; Lam and Carroll, 1983a). This heterogeneity can be organized such that similar variant repeats are located on the same chromosome (Lee and Singer, 1982; Appels and Dvorak, 1982).

In the present study we have examined highly repeated sequences from the grasshopper Caledia captiva. This species was chosen because of the presence of extreme variability in the amount and cytological distribution of heterochromatin among the C. captiva taxa (Shaw et al., 1980). In addition, preliminary molecular analyses revealed sequence variation (this paper and Arnold, unpub. data) within a particular highly repeated sequence family.

## Material and Methods.

### C. captiva Samples

For clarity, the taxa and populations examined in the present study have been expressed by numerals and letters, respectively. Initially, on the basis of chromosomal variation, C. captiva was divided into four races (Moran and Shaw, 1977) that were designated "South-east Australian" (S.E.A.), "Moreton", "Torresian" and "Daintree". However, after laboratory hybridization studies, demographic surveys and allozyme analyses it was demonstrated, using the criteria of levels of reproductive isolation and genic divergence, that the S.E.A. and Moreton races should be placed within the same subspecies, the Torresian "race" was designated a second subspecies and the Daintree "race" was given sibling species status (Shaw et al., 1980; Daly et al., 1981). In the present paper the S.E.A. and Moreton races, the



Torresian subspecies and the Daintree sibling species are designated "1", "2", "3" and "4", respectively. We will refer to these different taxonomic groups as taxa. The following populations were used in the in situ and/or sequencing experiments: Gundaroo, Araluen and Lakes Entrance (1A, 1B and 1C, respectively); Scrubby Creek and Peregrine (2A and 2B, respectively); Bongmuller, Insulator Creek and Morehead (3A, 3B and 3C, respectively); and Daintree (4A).

#### DNA Isolation

Samples of DNA were isolated from individual, eviscerated grasshoppers and from pooled samples containing 10 - 40 individuals using the techniques of Appels and Dvorak (1982). Yields of DNA from individual grasshoppers ranged from approximately 80 µg to 1000 µg, depending on size.

#### Cloning of the Highly Repetitive Sequence from Caledonia

DNA renaturing at a  $Cot < 0.02$  (mol x sec)/l was isolated from population 1A and used to synthesize radioactive probes ( $^{32}P$  c-DNA or  $^3H$  c-RNA) by standard procedures. A Taq I digest of total DNA from the 1A sample was probed with this DNA and a fragment consisting of approximately 150 base pairs was shown to be a subset of the  $Cot < 0.02$  fraction (Fig. 1). This fragment was isolated from agarose gels by electroelution. These Taq I fragments were then inserted into the Cla I site of pBR322 and used to transform E. coli ECR291; 2 out of 6 plasmids showed significant hybridization to the  $Cot 0.02$  fraction. Colonies

Fig. 1. Autoradiograph from "Southern" hybridization of Taq I  
digested total population 1A DNA to a *nick translated* 168 bp  
repeat from population 1A.



**168 bp**



were selected for inability to grow in the presence of 15 µg/ml of tetracycline while retaining their resistance to 200 µg/ml of ampicillin. A sample of clones sensitive to tetracycline was selected for individual plasmid preparations. Plasmid DNA from these clones was digested with Taq I, electrophoresed in a 2% agarose gel and then transferred to Gene Screen (New England Nuclear). The filter was exposed to a  $^{32}\text{P}$ -nick-translated "Cot 0.02" probe for 16-20 hr in 50% formamide, 0.5% SDS, 3 X SSC and 5 X Denhardt solution. The filter was washed for at least 4 hr in 2 X SSC/0.1% SDS, dried, covered with pre-exposed X-ray film and placed at  $-70^{\circ}\text{C}$  (Laskey, 1980), to screen for hybridizing clones.

#### Restriction Endonuclease Analysis

The following restriction enzymes were used in an initial attempt to map the cloned insert: Bam HI, Eco RI, Hind III, Pst I, Pvu I, Acc I, Ava II, Xma III, Rsa I, Hae III, Hha I, Sau 3A1 and Cla I. Rsa I was the only enzyme which demonstrated the ability to digest the sequence.

#### In situ Hybridization

Mitotic cells were prepared on air dried slides from 8-10 day old embryos (Webb, 1976) from population 1B. The chromosomal location of sequences homologous to the "150 bp" cloned sequence was demonstrated by the hybridization of  $^3\text{H}$ -cRNA to these preparations following the procedure of Appels et al. (1978).

#### Saturation Hybridization

Saturation hybridization experiments to determine copy number of the repetitive sequence in population 1A followed the

procedure of Bishop et al. (1969).

#### M13 Cloning and Dideoxy Sequencing of "Taq Family" Repeats

A 100 µg sample of the plasmid pSEA1 was digested with the restriction enzyme Taq I. This sample was electrophoresed in a 2% agarose gel and the band corresponding to the C. *captiva* DNA was cut out of the gel and subsequently electroeluted at 40 mA for 3-4 hr. This DNA was then ligated into the Acc I site of M13 mp8 (Messing, 1981).

Taq family repeats were isolated from two individuals from population 1A and from pooled grasshoppers for populations 1C, 2A, 2B, 3A, 3B, 3C and 4A. The isolation of the Taq family repeats from these samples was identical to the isolation of the repeat that provided the pSEA1 clone (see "Cloning of the Highly Repetitive Sequence from Caledia"). These sequences were cloned directly into the Acc I site of M13 mp8, mp9 or mp10 (Messing, 1981; Messing and Vieira, 1982), without first cloning into the Cla I site of pBR322. This facilitated the screening of larger numbers of cloned sequences in order to isolate the Taq family representatives from taxa 3 and 4 which contained 20% and 2%, respectively, of the number of copies in population 1A (Arnold and Shaw, in prep.).

The cloned repeats were sequenced using the method of Sanger et al. (1977); part of an autoradiogram from a representative sequencing gel is shown in Figure 2. The repeats were defined by having the Taq I recognition sequence (TCGA) at both their 5' and 3' ends; therefore, we have included two of the four bases at the

Fig. 2. Portion of X-ray from representative sequencing gel showing positions 35-84 from the 168 bp sequence. From left to right the lanes contain Adenine, Cytosine, Guanine and Thymidine reaction mixtures, respectively.



c g t c t a t t t c g t a c c t a t t t c t a t c



a g t a t c t t t c c a t g a t t a t a t t



5' (GA) and 3' (TC) ends as part of each repeat. Because the fragment is only 168 bp long, its sequence can be readily assessed from a single, 8% sequencing gel 0.75 meters long. The pSEA1 sequence was determined by carrying out the sequencing reactions on four independent M13 clones, three of which contained the same strand and one the opposite strand. No ambiguous positions occurred in the sequence. In particular, the 168 bp sequence analyzed in this study appeared to be free of regions that cause the "pile-ups" of the type discussed by Bankier and Barrell (1983). The M13 clones from the population analyses were examined by replicate sequencing reactions and once again no ambiguous positions occurred in the sequences. For the analysis of the 1A individuals, replicate sequencing reactions were carried out for eight out of the total of 39 clones examined. A consensus sequence was determined for each of the individuals analyzed by scoring the most common nucleotide at each position among the clones.

#### End-Labeling and DNA Sequencing with Base-Specific Chemical Cleavages

The 168 bp "monomers" of uncloned Taq family repeats were isolated by Taq I digestion, agarose gel electrophoresis and electroelution (see "Cloning of the Highly Repetitive Sequence from Caledonia") and end-labelled by the "fill-in" reaction (Goodman, 1979). These fragments were then digested with Rsa I and subjected to the DNA sequencing reactions described by Maxam and Gilbert (1980).

## Results.

### 1. Highly repetitive DNA in *C. captiva*

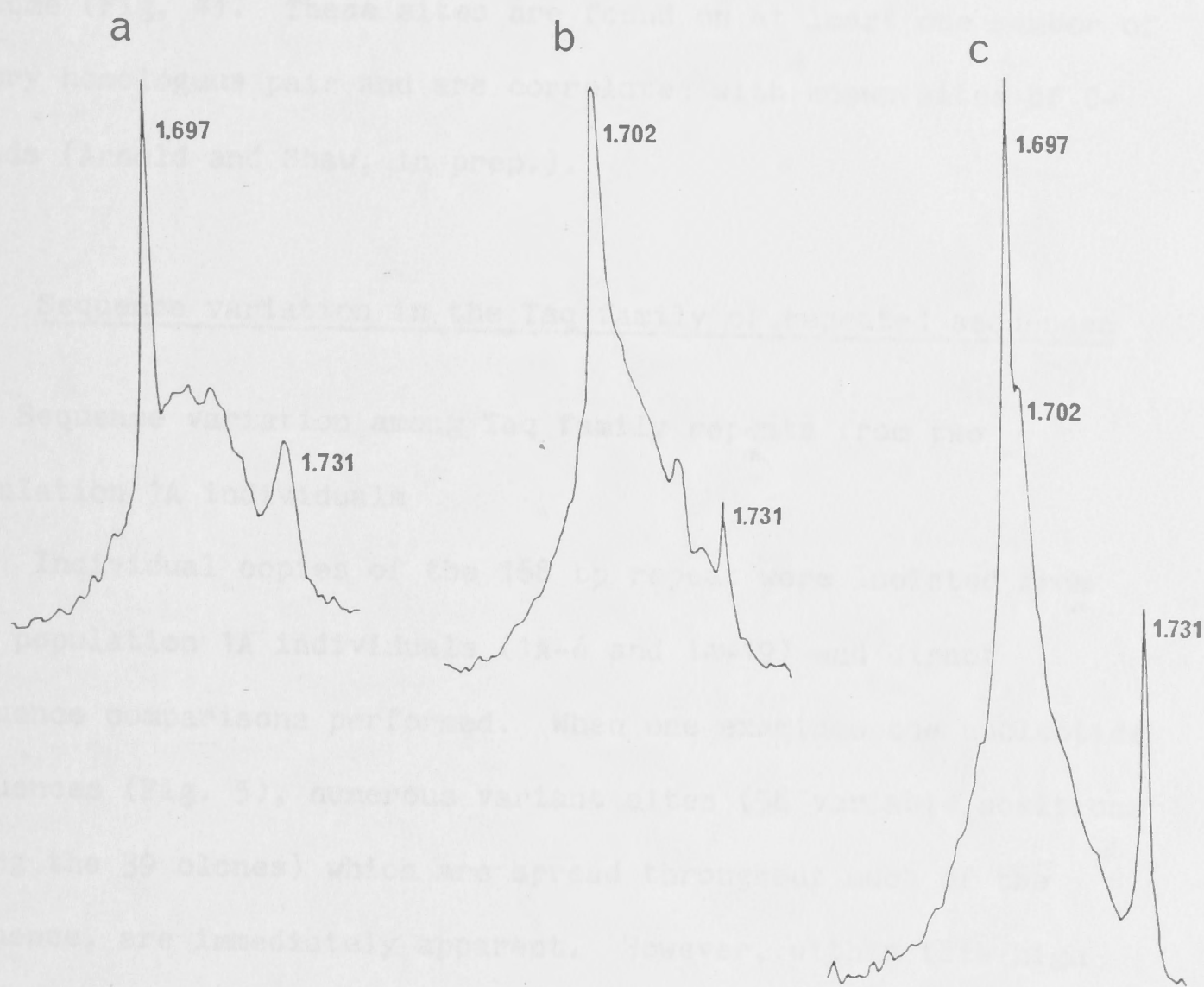
An initial characterization of highly repetitive DNA was carried out by examining the fraction recovered when total DNA was sheared to an average length of 500 bp, denatured and then renatured to a Cot of 0.02; 1-2% of the DNA from populations 2A and 4A of *C. captiva* was recovered in such a Cot 0.02 fraction. Buoyant density analyses in CsCl showed the presence of a hyper-sharp peak in both these populations as well as a broad range of the other DNA species (Fig. 3). The hyper-sharp peaks were of particular interest in this study because they indicated the presence of a major class of repetitive DNA that upon renaturation could form high molecular weight aggregates, a characteristic of simple repetitive sequences. Furthermore, the fact that the hyper-sharp peaks from populations 2A and 4A were demonstrably different when the two hyper-sharp peaks were mixed and banded together (Fig. 3) indicated that shifts in the base composition of the highly repetitive DNAs had occurred in the differentiation of these taxa. Sequences from population 2A are the subject of this paper while sequences from population 4A are presently being characterized (Arnold and Shaw, in prep.).

### 2. The Taq family repetitive sequence

Saturation hybridization experiments for population 1A from *C. captiva* indicated that approximately 0.2% of the total DNA, or  $1.7 \times 10^5$  copies of the Taq family sequence were present in the haploid genome.



Fig. 3. CsCl buoyant density profiles for a) population 2A rapidly reannealing DNA, b) population 4A rapidly reannealing DNA, and c) a mixture of the hypersharp peaks from populations 2A and 4A rapidly reannealing DNA.



The chromosomal distribution of the Taq I sequence was examined by in situ hybridization using a  $^3\text{H}$  c-RNA probe synthesized from the pSEA1 insert (the nucleotide sequence of this DNA is also shown in Fig. 4). Numerous interstitial and telomeric sites of hybridization are present in the population 1B genome (Fig. 4). These sites are found on at least one member of every homologous pair and are correlated with known sites of C-bands (Arnold and Shaw, in prep.).

### 3. Sequence variation in the Taq family of repeated sequences

#### (i) Sequence variation among Taq family repeats from two population 1A individuals

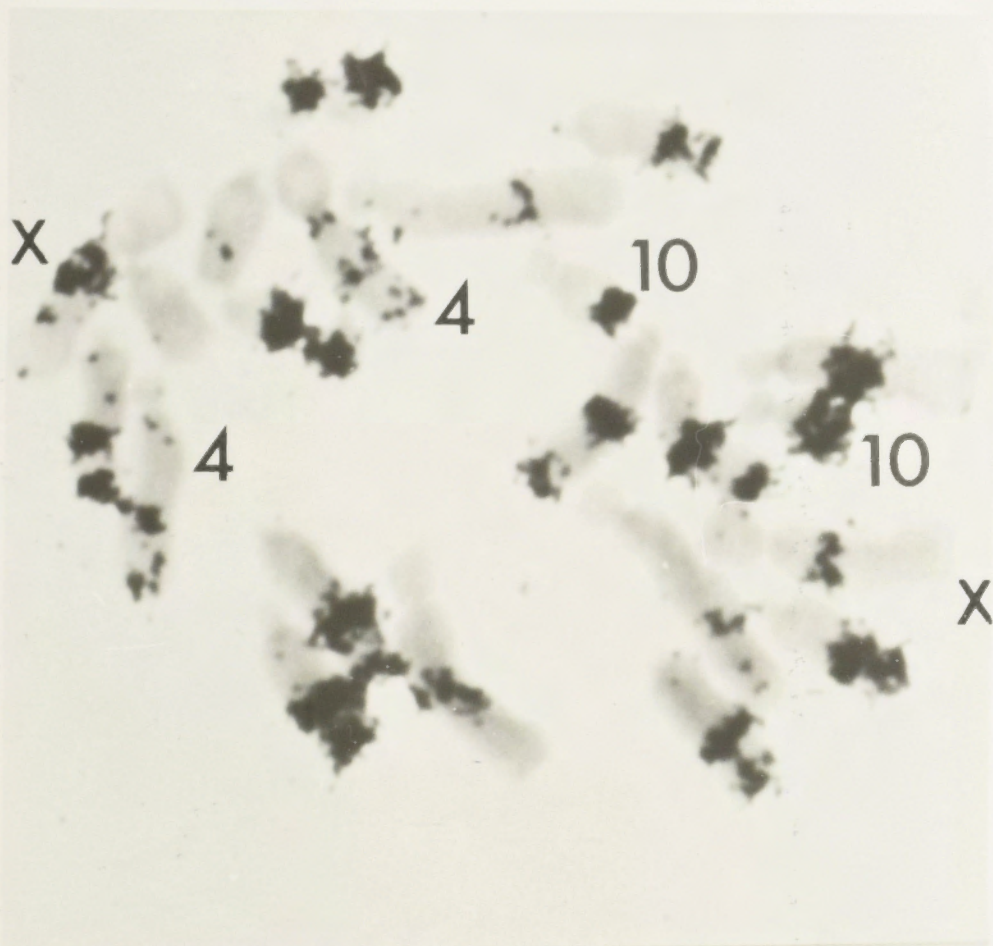
Individual copies of the 168 bp repeat were isolated from two population 1A individuals (1A-6 and 1A-19) and direct sequence comparisons performed. When one examines the nucleotide sequences (Fig. 5), numerous variant sites (58 variable positions among the 39 clones) which are spread throughout much of the sequence, are immediately apparent. However, within this high level of variation there are stretches of unaltered sequence, the longest being 18 bp (positions 67-84). This 18 bp conserved region is found in each of the 39 clones (Fig. 5), is 72% A+T and contains two direct repeats (CATT) separated by an adenine.

When each of the clones is compared to the consensus sequence determined for the individual from which they were derived (Fig. 5), a wide range of variation becomes apparent. The cloned repeats from 1A-6 and 1A-19 differ from their consensus sequence by 1.0% to 6.0% (1-10 bp) and 1.0% to 9.0% (1-



Fig. 4. In situ hybridization of  $^3\text{H}$  c-RNA from pSEA1 insert onto a population 1B chromosome complement. Note the polymorphisms in amount and/or position of hybridization on chromosomes 4, 10 and X. The nucleotide sequence for the 168 bp pSEA1 insert utilized in this experiment is also illustrated.





GAAGTCCATTCTAATTACTGATGCCTGT  
GGTTTGCTTTATTACTTATTTTAACGCT  
CGATTCGCCTTTTTTTCATTACATTGGCA  
TTAAATGGAGTGTATATTATGAAAATTG  
AGAGATAAAAATACTTAGAAGGGCTGCC  
ACTGGAACAGTGGGTCACTGCATAGTTC



**Fig. 5.** Repeats from two individuals, 1A-6 (sequences 1-19) and 1A-19 (sequences 20-39), and their respective consensus sequences (40 and 41). The 18 bp conserved region is indicated by bold-face type. Sequences 1, 20 and 40 are written in their entirety. For the other sequences, 2-19, 21-39 and 41 (the 1A-19 consensus sequence) only those nucleotides are shown that differ from the complete sequence immediately above them. Boxes (□) indicate the absence of a base (A = adenine, C = cytosine, G = guanine, T = thymine). Arrows (▼) located above the 1A-6 consensus sequence (40) denote positions at which substitutions, additions or deletions occur among the 39 cloned repeats.



[illegible]



15 bp), respectively. Comparison of the two consensus sequences reveals different bases at positions 3, 6, 16 and 60 that correspond to the highly variable positions in the cloned repeats (Fig. 5). Pairwise comparisons were carried out among the clones from each of the 1A individuals and the results of these comparisons are given in Table 1. From 0-10% variation in base pair sequence is present within these two individuals with a mean of 4% and 5% for 1A-19 and 1A-6, respectively.

(ii) An 18 bp sequence from within the Taq family repeat is conserved in C. captiva taxa.

Figure 6 presents sequence data for clones from eight C. captiva populations representing the races, subspecies and sibling species of C. captiva. Comparing each of the repeats shown in Figure 6 (with the exception of the population 4A sequences) with the consensus sequence based on the 39 clones from 1A-6 and 1A-19 indicated that the divergence ranged from 1.2% to 7.1%. In addition, pairwise comparisons were made among all of the 11 clones from these taxa. From this analysis values of 6%, 5%, 4% and 6% base pair divergence were detected for the intrapopulational, interpopulational, interracial and intersubspecific comparisons, respectively (Table 1). Much of this variability was located at the 58 positions previously identified in the 1A individuals. An additional 14 variable positions were, however, identified by the inter-taxon comparisons. Thus, a total of 72 variable positions were identified among the 50 clones analyzed from populations 1A, 1C, 2A, 2B, 3A, 3B and 3C. In spite of this high level of

Table 1. Mean base pair change and the range of base pair change at each of the intraspecific levels of divergence. Mean percent bp divergence is the number of nucleotide positions that do not match (including gaps), times 100, divided by the length of the consensus sequence (ie. 168 bp). The following pairwise comparisons were made: between the respective clones from individuals 1A-6 and 1A-19 (intraindividual values); between the respective clones from populations 2A, 2B, 3A and 3C (intrapopulation values); between the clones from populations 1A and 1C, 2A and 2B and 3A, 3B and 3C (interpopulation values); between the clones from populations 1A and 1C with the clones from populations 2A and 2B (interracial values); and between the clones from populations 1A, 1C, 2A and 2B with the clones from populations 3A, 3B and 3C (intersubspecific).



<u>Comparison</u>	<u>Mean Percent bp Divergence</u>	<u>Range of Percent bp Divergence</u>
Intraindividual	5.0	0-10
Intrapopulation	6.0	4-8
Interpopulation	5.0	2-9
Interracial	4.0	2-7
Intersubspecific	6.0	1-10

Fig. 6. Nucleotide sequences for cloned repeats from populations 1A(1), 1C(2), 2A(3,4), 2B(5,6), 3A(7,8), 3B(9), 3C(10,11) and 4A(12-14) along with the consensus sequences for taxa 1(15), 2(16) and 3(17). Bold-face type indicates the 18 bp conserved region in repeats 1-14 and the consensus sequences (15-17). Sequences 1, 12, 13 and the taxon 1 consensus sequence, 15, are written in their entirety. The format and symbols are as in Fig. 5. Arrows (▼) located above the taxon 1 consensus sequence (15) denote positions at which substitutions, additions or deletions occur among any of the sequences 1-11 in this figure or sequences 1-39 in Figure 5.



1	GAAGTGCATTCTAATTACTGATGCCTGTGGTTTGCTTTATTACTTATTTTAACGCTCGATTGCGCTTTTTTCATTACATTGGCATTAAATGGAGTGTATATTATGAAAATTGAGAGATAAAAAATACTTAGAAGGGCTGCCACTGGAACAGTGGGTCACTGCATAGTTC
2	T
3	T
4	T
5	T
6	T
7	T
8	T
9	T
10	T
11	T
12	GATGTGCATTCTATGTACTATTGCCTGTGGTTTGCTTTATTACTTATTTTAACGCTCAACTCGCCTTTTTTCATTACATTGGCATTAAATGGAGTGTATATTATGAAAATTGAGAGATAAAAAATACTTAGAAGGGCTGCCACTGGAACAGTGGGTCACTGCATAGTTC
13	GAAGGATCCTTTCTCCTAATCGGCTCGGTAATCTGCCCTTCTTAAGTTTTGCGTAGCATTGGATGTGGGTGCTATACCTATACAAAGTTTCATAAGGGGCGAGCTGACGGGACTGTAAATACCTCGGCTTGAATATAGCAGGTAAAGGTACCGCTGCGCGACCGCGGCATTTTTACCGCTTTTC
14	T
15	GAAGTCCATTCTATGTACTGATGCCTGTGGTTTGCTTTATTACTTATTTTAACGCTCAACTCGCCTTTTTTCATTACATTGGCATTAAATGGAGTGTATATTATGAAAATTGAGAGATAAAAAATACTTAGAAGGGCTGCCACTGGAACAGTGGGTCACTGCATAGTTC
16	T
17	T

1 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170



variability, the 18 bp region conserved in individuals 1A-6 and 1A-19 remained invariant; the occurrence of this stretch departed from random expectation ( $P < 0.01$ ) when tested by the method suggested by Brown and Clegg (1983).

In addition to the analysis of the above clones, three repeat units were isolated and sequenced from population 4A. The copy number of the Taq family repeat units in this population is only 2% of that found in population 1A (Arnold and Shaw, in prep.). It is clear from inspecting these sequences (Fig. 6) that Taq family repeats in this population are highly heterogeneous. <sup>1</sup>One repeat has only four base pair differences when compared to the consensus sequence from the population 1A individuals (Fig. 6). This sequence also contains the 18 bp conserved region present in all of the preceding C. captiva repeats. In contrast, the two remaining sequences (Clones 13 and 14; Fig. 6) are highly divergent in comparison to the other C. captiva repeats (Fig. 6). However, the 18 bp region discussed above is still recognizable even within these two divergent 4A sequences, although it is displaced 23 bp toward the 5' end of the repeat (Fig. 6). In this region of the two divergent 4A clones a total of 10 out of the 18 bp are identical to the consensus sequence from the 1A individuals.

<sup>1</sup>Subsequent experiments have suggested that this single clone may be a contaminant; this is presently under investigation.



(iii) The consensus sequence from 39 clones of the taxon 1 repeat is representative of the Taq family repetitive units

To relate the consensus sequence derived from cloned members of the Taq family to the entire population of Taq family repeats, we utilized two different assays.

One assay is based on the observation that an Rsa I site is located 16 bp from the 5' end of the repeat unit and occurs in 56% of the clones from population 1A. To assay the frequency of this site in the entire population of Taq family repeats the DNA from approximately 40 individuals from population 1A was digested with Taq I and the band containing the 168 bp repeats was recovered from a preparative agarose gel. The DNA was end-labeled using the "fill-in" reaction (see Material and Methods) and digested with Rsa I. Electrophoresis of the digest on a 6% polyacrylamide gel resolved three fragments (168 bp, 152 bp, 16 bp). Densitometric analysis of an autoradiogram of the gel indicated that 38% of the Taq family repeats contained the Rsa I site. The less frequent occurrence of the Rsa I site would be predicted on the basis that the band of DNA that corresponded to the 168bp repeats (see above) would also include other DNA species. Therefore, this frequency is consistent with the interpretation that the cloned sequences are not a highly selected subpopulation of repeat units.

The second assay utilized the above 152 bp end-labelled Rsa I fragment in a Maxam and Gilbert sequencing protocol (see Material and Methods). Preliminary sequence data for positions 72-116 (using the consensus sequence from taxon 1 as a reference) were obtained. This region was of particular interest because it



included much of the highly conserved 18 bp region discussed earlier. Only three ambiguous positions (72, 77, 83) were detected, indicating once again that the consensus sequence presented in Figure 5 is a useful approximation of the consensus sequence for the population of Taq family repeat units in the genome.

### Discussion.

This study has revealed extensive variation in nucleotide sequences at each of the levels of comparison from intraindividual to intersubspecific. Similar amounts of variation have also been reported by Strachan et al. (1985) between cloned repeats of the 360 bp family isolated from pooled samples of Drosophila from a single species. In addition, a number of clones from several Drosophila species were compared and in all but one of these comparisons there was a tenfold increase in the amount of divergence among the different cloned repeats (Strachan et al., 1985). This observation of extensive divergence between species with relatively low levels of within species variation in both multigene families and highly repeated DNA has been termed concerted evolution (Zimmer et al., 1980; Dover, 1982; Arnheim, 1983). Although there have been numerous documentations of this phenomenon between species, no study has extensively investigated the question of how variation is partitioned among populations that have not yet diverged to the level of biological species. The data presented in the Caledia study are directly relevant to the question of whether or not



sequence variation in highly repeated DNA is correlated with levels of divergence below that of the species.

Shaw et al. (1980) and Coates and Shaw (1984) have described the patterns of reproductive isolation within and among the C. captiva taxa. The level of genetic compatibility between the different populations, in crosses involving different races among backcross and F2 generations, ranged from no reduction in viability to a 50% and 100% reduction in these hybrid types, respectively, in crosses between individuals from different subspecies. Allozyme data for each of the four taxa in the C. captiva species complex have shown that protein electrophoretic divergence is correlated with the levels of reproductive isolation (Daly et al., 1981). Whatever mechanisms cause sequence divergence in the highly repeated Taq family, they are not directly correlated with taxonomic divergence below the species level since the mean nucleotide divergence between repeats from the same individual is similar to that found in comparisons between subspecies ( <sup>5</sup> % and 6%, respectively).

In contrast to the high levels of both intraspecific variation and interspecific divergence among highly repeated DNA families, numerous studies have detected significant similarities between tandemly repeated sequences from related animal taxa (see Singer, 1982 for review). The estimates of sequence similarity were derived indirectly from melting point determinations, positions of restriction endonuclease cleavage sites and the length of the basic repeating unit. For example, Strachan et al. (1982) used each of these criteria in their analysis of the evolution of several tandemly repeated DNA families from



Drosophila species. Two repeated DNA families (360 bp and 500 bp) are highly similar in both D. simulans and D. mauritiana whereas significant differences were observed between D. yakuba and D. teissieri. Lam and Carroll (1983a) have presented data on a 741 bp repeat from Xenopus laevis which, on the basis of restriction endonuclease cleavage sites, indicate that variation among the repeats from this family is extensive. No significant similarity was found between this sequence from X. laevis and the genomes of two other Xenopus species. In contrast, a 388 bp repeated family did share considerable similarity with sequences from two other Xenopus species (Lam and Carroll, 1983b). Significant similarities between highly repeated sequences have also been detected for Mus species and between Ovis and Bos highly repeated sequences. Pietras et al. (1983) identified a minor satellite sequence in a Mus musculus repetitive DNA library, for which the repetition frequency was 5 - 10% of that of the major satellite. Comparison of this minor satellite with the major satellite from M. musculus resolved a 29 bp region of extensive similarity between these two highly repeated families. Furthermore, this study demonstrated that sequences derived from Mus spretus are also related to the minor satellite from M. musculus. Finally, a recent analysis of sheep highly repeated DNA has shown that three distinct repeat families (370 bp, 435 bp and 800 bp) can be resolved by restriction endonuclease digestion with Eco RI (Novak, 1984). Each of these repeat families <sup>is</sup> related to one another on the basis of their nucleotide sequences and a comparison of the 435 bp repeat to the bovine 1.715



satellite sequence revealed more than 50% identity between these two sequences (Novak, 1984).

These observations of conserved highly repeated DNA sequences among both closely and more distantly related taxa <sup>have</sup> led to the suggestion that there may be functional constraints upon the nucleotide composition of these sequences. It has, however, been pointed out by both Ohta and Dover (1983) and Miklos (1985) that this conservation does not necessarily imply that natural selection is acting or has acted upon these sequences. The homogeneity of repeats between taxa can also be explained on the basis of recent origin, in which case an insufficient amount of time has elapsed for the accumulation of mutations, chance fixation of certain nucleotide sequences and/or a mechanism of biased conversion that leads to the fixation of a variant sequence (Arnheim, 1983; Ohta and Dover, 1984).

An examination of the C. *captiva* Taq family repeats reveals that the variation in nucleotide sequence, although encompassing approximately half of the base positions, is nonrandom. The obvious plasticity at some sites but not at others seems to indicate that there are limitations upon which of the sites can change at high frequency. Particularly striking is the occurrence of an 18 bp region that is unchanged in all but two of the clones. These two clones isolated from the sibling species (population 4A) are homologous to the 168 bp repeat both on the basis of filter hybridization (Arnold, unpub. data) and direct nucleotide examination, albeit at a much reduced level in comparison to the intraspecific values. This divergence has resulted in the shifting of the base sequence corresponding to

the 18 bp conserved region. Whether this is a consequence of mutations that have produced a new Taq I site or a rearrangement that has transposed this 18 bp region cannot be determined.

A similar pattern of conservation was detected by Dennis and Peacock (1984) when they sequenced the 180 bp repeat derived from the maize knob heterochromatin. A 27 bp sequence was found to be unvaried among the 19 clones isolated from maize, teosinte and Tripsacum. The presence of a highly conserved region, within a sequence which otherwise shows extensive variation, suggests the presence of a mechanism that prevents the accumulation of mutations at these base pair positions. One mechanism that could generate such a nonrandom pattern of change is biased gene conversion (Dover, 1982) with the "domain of conversion" being less than the repeat length of the Taq family sequence. One result of this process would be the generation of repeats with a mosaic pattern of change, with different regions of the sequence demonstrating different amounts of variation, while the same region in different repeats would be tending toward homogeneity (Ohta and Dover, 1984).

Clearly, many possibilities exist to explain the nonrandom pattern of divergence observed in Caledia, but it is interesting that the conserved 18 bp region within the Taq family repeats represents part of a two-fold symmetry (Fig. 7) of the type found in DNA sequences that bind proteins (Ptashne et al., 1980; Siebenlist et al., 1980; Borgmeyer et al., 1984). Previously, Hsieh and Brutlag (1979) reported the presence of two-fold symmetry in a Drosophila heterochromatic sequence (the 360 bp



Fig. 7. Nucleotide sequence at positions 60-115 from the consensus sequence for population 1A clones. Bold-face type indicates the 18 bp conserved region. Areas of dyad symmetry are shown within the boxed regions.

CTCGCCTTTTTTTCATTACATTGGCATTAAATGGAGTGTATATTATGAAAATTGAGA  
GAGCGGAAAAAGTAATGTAACCGTAATTTACCTCACATATAATACTTTTAACTCT

60 70 80 90 100 110



family) that was shown to bind preferentially to an embryo specific protein. In addition, highly repeated DNA (alphoid sequences) from primate species also demonstrate regions of dyad symmetry (see Miklos, 1985 for review). However, in both of these examples there is no apparent sequence conservation in the areas of two-fold symmetry in comparison to other regions of the repeat. For example, the area of dyad symmetry in the 360 bp family from Drosophila was found to be heterogeneous even among repeats from a single laboratory strain (Miklos and Gill, 1981).

The confirmation of a function for the C. captiva Taq family of repeats requires first, the demonstration of a specific interaction of this sequence with a protein component, and second, the identification of the role of such an interaction. However, the conservation of a portion of this repeated element in conjunction with the unique structure of this conserved region is suggestive of such a cellular interaction.

Acknowledgements. We wish to thank Dr. David Coates for his assistance with field collecting and Frances Arnold for her typing numerous drafts of the manuscript. We also thank Gary Brown and Maureen Whitaker for their assistance with the illustrations and photography, respectively. One of us (M.A.) is grateful to the Division of Plant Industry, CSIRO for providing laboratory facilities for some of the experiments performed. M.A. was supported by an Australian National University Ph.D. scholarship. Excellent technical advice was given by P. Wilkinson, L. Moran, F. Clay and B. Clarke. We are grateful to

B. John and two anonymous reviewers for their helpful comments on the manuscript.

#### Literature Cited

- Appels, R., and J. Dvorak. 1982. The wheat ribosomal DNA spacer region : its structure and variation in populations and among species. *Theor. Appl. Genet.* 63: 337-348.
- Appels, R., C. Driscoll, and W.J. Peacock. 1978. Heterochromatin and highly repeated DNA sequences in rye (*Secale cereale*). *Chromosoma* 70: 67-89.
- Appels, R., and W.J. Peacock. 1978. The arrangement and evolution of highly repeated (satellite) DNA sequences with special reference to *Drosophila*. *Int. Rev. Cytol. Suppl.* 8: 69-124.
- Arnheim, N. 1983. Concerted evolution of multigene families. Pp. 38-61 in M. Nei and R.K. Koehn, eds. *Evolution of genes and proteins*. Sinauer Associates, Inc., Sunderland, Mass.
- Bankier, A.T., and B.G. Barrell. 1983. Shotgun DNA sequencing. Pp. 1-34 in R.A. Flavell, ed. *Techniques in nucleic acid biochemistry*. Elsevier, Amsterdam.
- Bishop, J.O., F.W. Robertson, J.A. Burns, and M. Melli. 1969. Methods for the analysis of deoxyribonucleic acid-ribonucleic acid hybridization data. *Biochem. J.* 115: 361-370.



- Borgmeyer, U., J. Nowock, and A.E. Sipple. 1984. The TGGCA-binding protein: a eukaryotic nuclear protein recognizing a symmetrical sequence on double-stranded linear DNA. N.A.R. 12: 4295-4311.
- Brown, A.H.D., and M.T. Clegg. 1983. Analysis of variation in related DNA sequences. Pp. 107-132. in B.S. Weir, ed. Statistical analysis of DNA sequence data. Marcel Dekker, Inc., New York.
- Coates, D.J., and D.D. Shaw. 1984. The chromosomal component of reproductive isolation in the grasshopper Caledia captiva III. Chaisma distribution patterns in a new chromosomal taxon. Heredity 53: 85-100.
- Daly, J.C., P. Wilkinson, and D.D. Shaw. 1981. Reproductive isolation in relation to allozymic and chromosomal differentiation in the grasshopper Caledia captiva. Evolution 35: 1164-1179.
- Dennis, E.S., and W.J. Peacock. 1984. Knob heterochromatin homology in maize and its relatives. J. Mol. Evol. 20: 341-350.
- Dover, G. 1982. Molecular drive : a cohesive mode of species evolution. Nature 299: 111-116.
- Goodman, H.M. 1979. Repair of overlapping DNA termini. Meth. Enzymol. 65: 63-64.
- Hsieh, T., and D.L. Brutlag. 1979. A protein that preferentially binds Drosophila satellite DNA. P.N.A.S. 76: 726-730.

- John, B., and G.L.G. Miklos. 1979. Functional aspects of satellite DNA and heterochromatin. *Int. Rev. Cytol.* 58: 1-114.
- Lam, B.S., and D. Carroll. 1983a. Tandemly repeated DNA sequences from Xenopus laevis. I. Studies on sequence organization and variation in satellite 1 DNA (741 base-pair repeat). *J. Mol. Biol.* 165: 567-585.
- Lam, B.S., and D. Carroll. 1983b. Tandemly repeated DNA sequences from Xenopus laevis. II. Dispersed clusters of a 388 base-pair repeating unit. *J. Mol. Biol.* 165: 587-597.
- Laskey, R.A. 1980. The use of intensifying screens or organic scintillators for visualizing radioactive molecules resolved by gel electrophoresis. *Meth. Enzymol.* 65: 363-371.
- Lee, N.H.T., and M.F. Singer. 1982. Structural organization of  $\alpha$ -satellite DNA in a single monkey chromosome. *J. Mol. Biol.* 161: 323-342.
- Maxam, A.M., and W. Gilbert. 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. *Meth. Enzymol.* 65: 499-560.
- Messing, J. 1981. M13 mp2 and derivatives: a molecular cloning system for DNA sequencing, strand specific hybridization, and in vitro mutagenesis. Pp. 143-153. in A.G. Walton, ed. *Proceedings of the third Cleveland symposium on macromolecules*. Elsevier, Amsterdam.
- Messing, J., and J. Vierira. 1982. A new pair of M13 vectors for selecting either strand of double-digest restriction fragments. *Gene* 19: 269-276.



- Miklos, G.L.G. 1985. Localised highly repetitive DNA sequences in vertebrate and invertebrate genomes. *Evol. Theory*, accepted.
- Miklos, G.L.G., and A.C. Gill. 1981. The DNA sequences of cloned complex satellite DNAs from hawaiian Drosophila and their bearing on satellite DNA sequence conservation. *Chromosoma* 82: 409-427.
- Moran, C., and D.D. Shaw. 1977. Population cytogenetics of the genus Caledia (Orthoptera:Acridinae) III. Chromosomal polymorphism, racial parapatry and introgression. *Chromosoma* 63: 181-204.
- Novak, U. 1984. Structure and properties of a highly repetitive DNA sequence in sheep. *N.A.R.* 12: 2343-2350.
- Ohta, T., and G.A. Dover. 1983. Population genetics of multigene families that are dispersed into two or more chromosomes. *P.N.A.S.* 80: 4079-4083.
- Ohta, T., and G.A. Dover. 1984. The cohesive population genetics of molecular drive. *Genetics* 108: 501-521.
- Pietras, D.F., K.L. Bennet, L.D. Siracusa, M. Woodworth-Gutai, V.M. Chapman, K.W. Gross, C. Kane-Haas, and N.D. Hastie. 1983. Construction of a small Mus musculus repetitive DNA library: identification of a new satellite sequence in Mus musculus. *N.A.R.* 11: 6965-6983.
- Ptashne, M., A. Jeffrey, A.D. Johnson, R. Maurer, B.J. Meyer, C.O. Pabo, T.M. Roberts, and R.T. Sauer. 1980. How the  $\lambda$  repressor and Cro work. *Cell* 19: 1-11.
- Sanger, F., S. Nicklen, and A.R. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *P.N.A.S.* 74: 5463-5467.

- Shaw, D.D., C. Moran, and P. Wilkinson. 1980. Chromosomal reorganization, geographic differentiation and the mechanism of speciation in the genus Caledia. Pp. 171-194 in R.L. Blackman, G.M. Hewitt, and M. Ashburner, eds. Insect Cytogenetics, Symp. Roy. Entomol. Soc. London No. 10. Blackwells, Oxford.
- Siebenlist, U., R.B. Simpson, and W. Gilbert. 1980. E. coli RNA polymerase interacts homologously with two different promoters. Cell 20: 269-281.
- Singer, M.F. 1982. Highly repeated sequences in mammalian genomes. Int. Rev. Cytol. 76: 67-112.
- Strachan, T., E. Coen, D. Webb, and G. Dover. 1982. Modes and rates of change of complex DNA families of Drosophila. J. Mol. Biol. 158: 37-54.
- Strachan, T., D. Webb, and G. Dover. 1985. Transition stages during sequence turnover in multiple-copy DNA families in Drosophila. Cell, accepted.
- Webb, G.C. 1976. Chromosome organisation in the Australian plague locust, Chortoicetes terminifera. I. Banding relationships of the normal and supernumerary chromosomes. Chromosoma 55: 229-246.
- Zimmer, E.A., S.L. Martin, S.M. Beverley, Y.W. Kan, and A.C. Wilson. 1980. Rapid duplication and loss of genes coding for the  $\alpha$  chains of hemoglobin. P.N.A.S. 77: 2158-2162.
- Wesley M. Brown, reviewing editor.



**CHAPTER 4:** The Heterochromatin of Grasshoppers from the Caledia  
captiva species complex II. Cytological Organization  
of Tandemly Repeated DNA Sequences

This chapter has been submitted for publication in Chromosoma,  
and is in the format required for this journal. The authors are  
M.L. Arnold and D.D. Shaw

**Abstract.** C-band variation between the Caledia taxa is extensive with numerous large interstitial and telomeric blocks of heterochromatin being present in the South-east Australian and Moreton taxa while the Torresian types possess small centromeric or telomeric C-bands. In situ hybridizations using  $^3\text{H}$  c-RNA from the 168 bp highly repeated sequence defined further variation between the C. captiva taxa. This sequence family is present in each of the interstitial and telomeric constitutive heterochromatic blocks in the South-east Australian and Moreton taxa. However, it is represented in only a fraction of the heterochromatic regions, defined by C-banding, within the three Torresian types. The 144 bp sequence family is restricted to the procentric blocks of heterochromatin of chromosomes 2-7, 9 and 10 in the Daintree taxon and does not share homology with the 168 bp sequence. This sequence is A-T rich and possesses a region of dyad symmetry. Quantitative measurements for the two sequence families revealed a wide range of copy numbers between the C. captiva taxa. The 168 bp family has approximately 150,000, 35,000 and 4,000 copies, respectively, in the South-east Australian/Moreton, Torresian and Daintree genomes. There are 2,000,000 and 100,000 copies of the 144 bp sequence in the Daintree and Papuan Torresian taxa, respectively. The distributional, quantitative and sequence characteristics of these repeat families imply that past amplification or introgression has played a major role in the evolution of these sequences. There is an overall negative correlation between the quantity of the 168 bp sequence and the levels of reproductive isolation and genic divergence between the various taxa. It is



possible that some of the reduction in the viability of the hybrid individuals is due to the quantitative changes in these sequences. Moreover, the quantitative and qualitative characteristics of highly repeated DNA families may play a role in the modulation of such essential cellular functions as cell cycle duration, nuclear organization and gene expression.

## INTRODUCTION

Studies of heterochromatin localization, which have included both C-banding and in situ hybridization of highly repeated DNA probes, have, in some cases, identified major regions of highly repeated DNA within the genome (Miklos and Nankivell, 1976; Appels, 1982). Constitutive heterochromatin from a limited number of species has been demonstrated to contain a composite of various families of highly repeated or satellite DNA. This type of organization has been documented for telomeric (rye: Appels, 1982), centromeric (calf: Kurnit et al., 1973) and interstitial (Drosophila: Peacock et al., 1977) heterochromatin. Molecular analyses of tandemly repeated DNA families have also shown that homologous sequences can occur on from a few (Drosophila: Peacock et al., 1974) to most (mouse: Pardue and Gall, 1970) of the chromosomes in a complement. In addition, highly repeated sequences have been compared in closely and more distantly related species in terms of their sequence organization and composition (Sutton and McCallum, 1972; Gall et al., 1974; Peacock et al., 1978; Novak, 1984) as well as their cytological distribution (Peacock et al., 1978; Cohen and Bowman, 1979).

Each of these studies has demonstrated that there are high levels of interspecific variation in respect of the cytological distribution and sequence organization of highly repeated DNA families. Furthermore, Strachan et al. (1985) and Trick and Dover (1984) have shown that intraspecific sequence variation can be an order of magnitude less than (Drosophila) or equivalent to (Tsetse fly) the interspecific value. Thus, there exists a wide range of variation in the sequence organization and cytological distribution of highly repeated DNA both within and between different species.

An examination of a highly repeated DNA family ("Taq" family) from the grasshopper Caledia captiva has provided the first detailed analysis of sequence variation within individuals, populations, races and subspecies of a single species complex (Arnold et al., 1985a). This study was carried out to directly assess how highly repeated DNA sequences evolve in relation to the accumulation of genetic differences leading to reproductive isolation. In addition, the structure and pattern of evolution of the repeats within the Taq family were used to test for any potential functional roles of these sequences.

In the present study we have examined the patterns of C-band variation present in seven different taxa from the genus Caledia (Shaw et al., 1976; 1980; this study) and have compared these with the cytological distribution of both a 168 bp repeat family which originated from the Gundaroo population of the South-east Australian taxon and a second highly repeated 144 bp DNA family. This latter family is also defined by Taq I recognition sites at the 5' and 3' termini and was isolated from the Daintree taxon.



The amounts of these two sequence families were also determined for each of the taxa. These data were then used to compare the organization of heterochromatin within taxa of the genus Caledia with specific reference to these two repeated DNA families.

#### MATERIALS AND METHODS

Individuals from the following taxa were used in the present study: South-east Australian (S.E.A.; Gundaroo, N.S.W. and Lakes Entrance, Vic.); Moreton (Peregian Beach, Qld.); Torresian ("Southern" Torresian = Bongmuller and Insulator Creek (Ck), Qld., "Northern" Torresian = East Normanby, Qld., "Papuan" Torresian = Morehead, P.N.G.); Daintree (Daintree, Qld.); and Caledia species nova 1 (Morehead P.N.G.).

Mitotic chromosome complements were prepared from whole, eight to ten day old embryos on air dried slides (Webb, 1976). The locations of C-band positive material and of sequences homologous to the 168 bp and 144 bp cloned repeats was identified using the techniques of Webb (1976) and Appels et al. (1978), respectively. Copy numbers for each of the highly repeated DNA sequences were calculated using the procedure of Bishop et al. (1969).

Total DNA was isolated from pooled samples (10 - 40 individuals) from each of the taxa of C. captiva, along with the sibling species (C. captiva Daintree) and Caledia species nova 1. Cloning of the 168 bp Taq family fragment isolated from the S.E.A. taxon has been described in detail by Arnold et al.

(1985a). Isolation of the 144 bp sequence followed similar procedures. Total Daintree DNA was digested with the restriction endonuclease Taq I. Following electrophoresis in a 2.0% agarose gel and EtBr staining, several bands ranging in size from 100 - 200 bp were visible under ultraviolet light. The digests were transferred to Gene Screen (New England Nuclear) using the procedure of Southern (1975). The DNA was fixed to the filter by baking at 80°C for two hr. Prehybridization was carried out for one hr in hybridization solution without the radioactive probe. The filter was hybridized to total Daintree DNA, which had been <sup>32</sup>P-nick-translated, to a Cot value of approximately 0.01 (mol x sec)/l. The hybridization solution consisted of 50% formamide, 0.5% SDS, 3 X SSC and 5 X Denhardt solution. Following hybridization the filter was washed for one hr in 2 X SSC/0.1% SDS at room temperature, with four changes of the wash solution. The filter was then covered with pre-exposed X-ray film and stored at -70°C (Laskey, 1980). The most intensely labelled band was cut from the agarose gel and the fragments were electroeluted. An aliquot of this sample was then ligated into pBR322 which had been linearized with the restriction enzyme Cla I. These recombinant plasmids were used to transform *E. coli* RRI cells. Positive clones were selected on the basis of their homology to the Cot 0.01 fraction of Daintree DNA. One such clone (pDA1) was used in the present study.

DNA (100 µg) from the plasmid pDA1 was digested with Taq I endonuclease. Following electrophoresis, the band corresponding to the Daintree highly repeated insert was removed and electroeluted. These sequences were ligated into Acc I digested



M13mp10 (Messing and Vierira, 1982) and the recombinant phage were used to transform E. coli JM 101.

The dideoxy chain termination sequencing reaction (Sanger et al., 1977) was used to determine the nucleotide sequence of the Daintree cloned repeat. Because this repeat is defined by Taq I recognition sites (TCGA) at the 5' and 3' ends we have included two of the four bases at the 5' (GA) and 3' (TC) termini. Both strands of this repeat were sequenced with no ambiguous results.

## RESULTS

### i) Gross karyotypic structure

The structure of the karyotype has previously been described for several C. captiva populations (Shaw, 1976; Shaw et al., 1980; Coates and Shaw, 1984). We summarize these descriptions in Table 1 along with data from a number of new populations, namely Araluen, Gundaroo, Insulator Ck and East Normanby. Each of the Torresian populations, as well as the S.E.A. population from Lakes Entrance, are characterized by acrocentric chromosomes. The Daintree and C. species nova 1 individuals possess telocentric karyotypes, whilst the Moreton and remaining S.E.A. populations have karyotypes with meta-, acro- and telocentric chromosomes. The Moreton and Lakes Entrance populations represent the endpoints of a continuous chromosomal cline in which intermediate populations show complex polymorphisms (Shaw et al., in prep.).

### ii) C-band heterochromatin

Shaw and his coworkers (Shaw et al., 1976; Shaw et al.,

Table 1. Distributional and quantitative variation for the C-band heterochromatin and highly repeated DNA in the taxa of Caledia. (N.B. The terms telomeric and centromeric are used to indicate the presence of heterochromatin located near or at these chromosomal structures.) N.D. = Not determined. Dashes (--) indicate that the repeat sequence is absent. S., N. and P. Torresian = Southern, Northern and Papuan Torresian, respectively.



TAXON	POPULATION	CENTROMERE POSITION	C-BAND HETEROCHROMATIN DISTRIBUTION (Numbers refer to chromosome pairs)			HIGHLY REPEATED DNA DISTRIBUTION		HIGHLY REPEATED DNA COPY NUMBER	
			CENTROMERIC	TELOMERIC	INTERSTITIAL	168 bp FAMILY	144 bp FAMILY	168 bp FAMILY	144 bp FAMILY
S.E.A.	Araluen	Polymorphic	1-12	4-8, 10-12	1-12	Same As Heterochromatin	--	N.D.	--
	Gundaroo	Polymorphic	1-12	4-8, 10-12	1-12	N.D.	--	1.5x10 <sup>5</sup>	--
	Lakes Entrance	Acrocentric	1-12	4-8, 10-12	1-8, 11	Same As Heterochromatin	--	N.D.*	--
Moreton	Peregrin	Metacentric	1-4, 6-12	4-8, 10-12	1-4, 6-9, 11, 12	Same As Heterochromatin	--	N.D.*	--
Torresian	Bongmuller (S. Torresian)	Acrocentric	1-12	10-12	10-11	Telomeres 10, 11, 12	--	3.5x10 <sup>4</sup>	--
	Insulator Ck (S. Torresian)	Acrocentric	1-12	10-12	10, 11	Telomeres 10, 11	--	N.D.	--
	E. Normanby (N. Torresian)	Acrocentric	1-12	2, 4, 10, 11	10, 11	Telomere 11	--	N.D.	--
	Morehead (P. Torresian)	Acrocentric	1-12	1, 8, 10-12	X, 4, 9, 10	Telomeres 6, 12	N.D.	N.D.**	1.0x10 <sup>5</sup>
Daintree	Daintree	Telocentric	1-12	12	10	N.D.	Procentric Heterochromatin	3.8x10 <sup>3</sup>	2.0x10 <sup>6</sup>
<u>C. species nova 1</u>		Telocentric	1-12	2, X, 5-12	4, 7, 11, 12	--	--	--	--

\* Preliminary data, along with the in situ results suggest that these populations have approximately the same amount of the 168 bp sequence as the Gundaroo samples.

\*\*The in situ data suggest that this population possesses an approximately equivalent amount of this sequence to that found in the Bongmuller sample.

1980; Coates and Shaw, 1984) have also defined a complex pattern of C-band heterochromatin variation within the C. captiva species group. Figure 1 (compare with Table 1) presents data for the previously described taxa along with results for Papuan Torresian, Northern Torresian and C. species nova 1 individuals. There is extensive variation in the distribution and amount of heterochromatin between the various taxa.

iii) In situ localization of highly repeated sequences

Figure 2(a-f) illustrates the variation in the cytological locations of the 168 bp (Fig. 2a-e) and 144 bp (Fig. 2f) sequences. The two S.E.A. individuals (Fig. 2a, b) and the Moreton animal (Fig. 2c) have numerous interstitial and telomeric sites of hybridization. At least in the Moreton taxa, these sites of hybridization correspond to the C-band positive regions (Arnold et al., 1985b). The distribution of the 168 bp sequence is also presented for the Southern Torresian and Papuan Torresian taxa (Fig. 2d, e). Southern Torresian individuals show hybridization to the telomeres of chromosomes 10, 11 and 12 (Fig. 2d and Arnold et al., 1985b). The Papuan Torresian karyotype illustrated in Figure 2e has the 168 bp sequence present on the telomeres of chromosomes 6 and 12, but not on those of 10 or 11. Two other Torresian populations (Insulator Ck and East Normanby) were also assayed in the in situ hybridization experiments involving the 168 bp sequence. The Insulator Ck population (Southern Torresian) exhibited hybridization in the telomeric regions of chromosomes 10 and 11, while the East Normanby individuals (Northern Torresian) had hybridization only at the

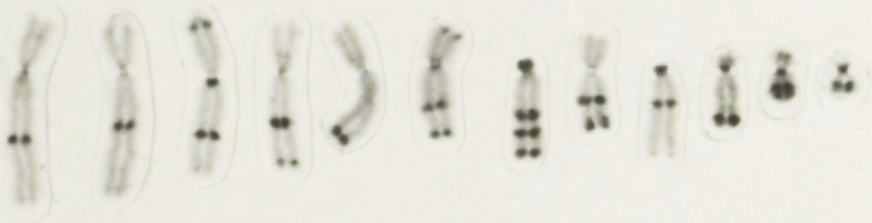


Fig. 1 C-banded haploid karyotypes of the Caledia taxa.

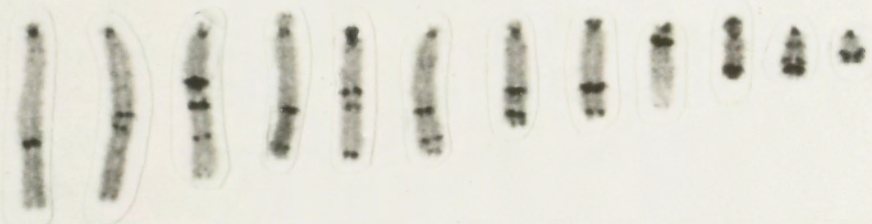


1 2 X 4 5 6 7 8 9 10 11 12

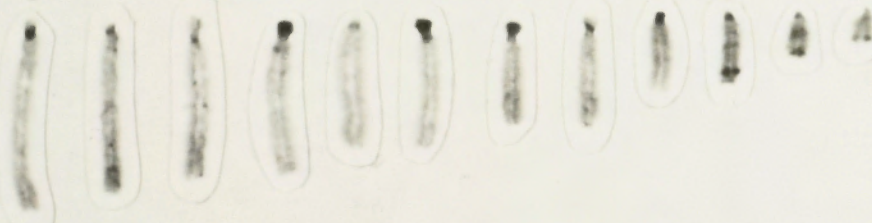
*C. captiva*  
(MORETON)



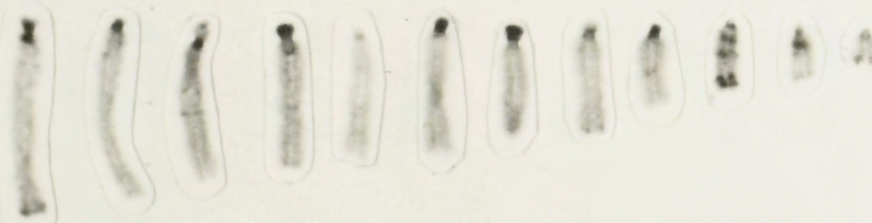
*C. captiva*  
(LAKES  
ENTRANCE)



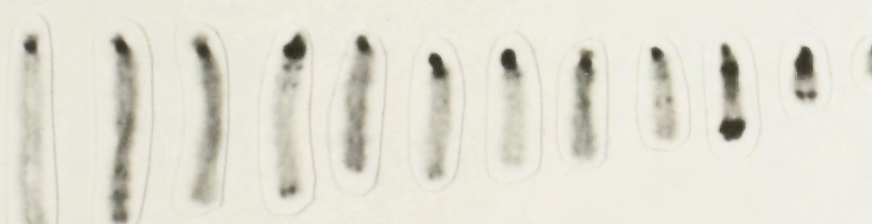
*C. captiva*  
(SOUTHERN  
TORRESIAN)



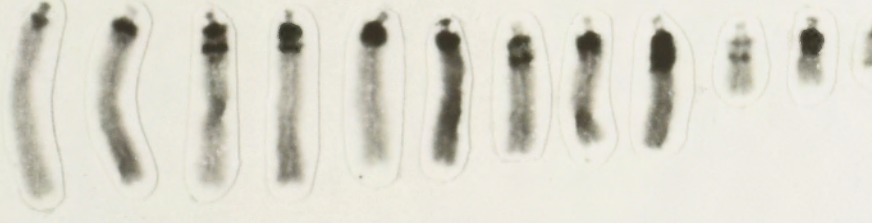
*C. captiva*  
(PAPUAN)



*C. captiva*  
(NORTHERN  
TORRESIAN)



*C. captiva*  
(DAINTREE)



*C. species*  
nova 1

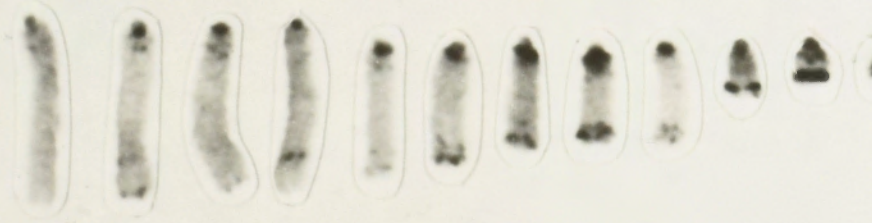
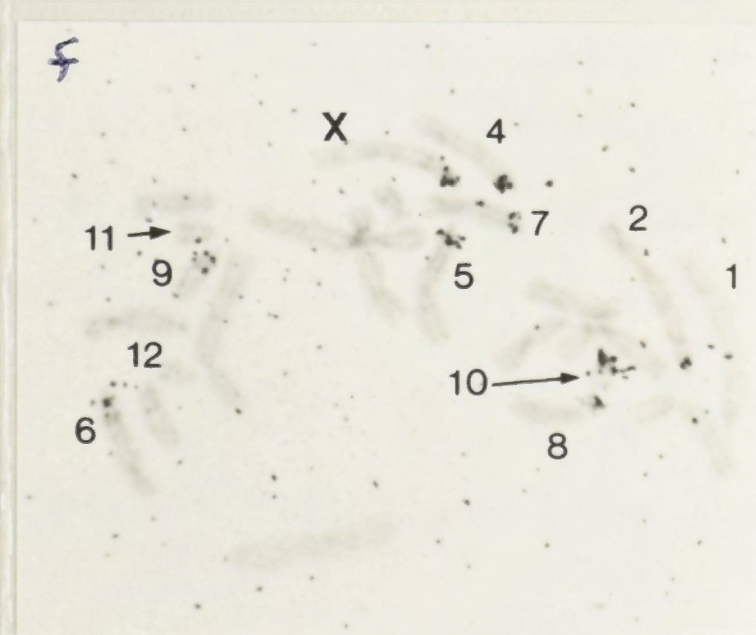
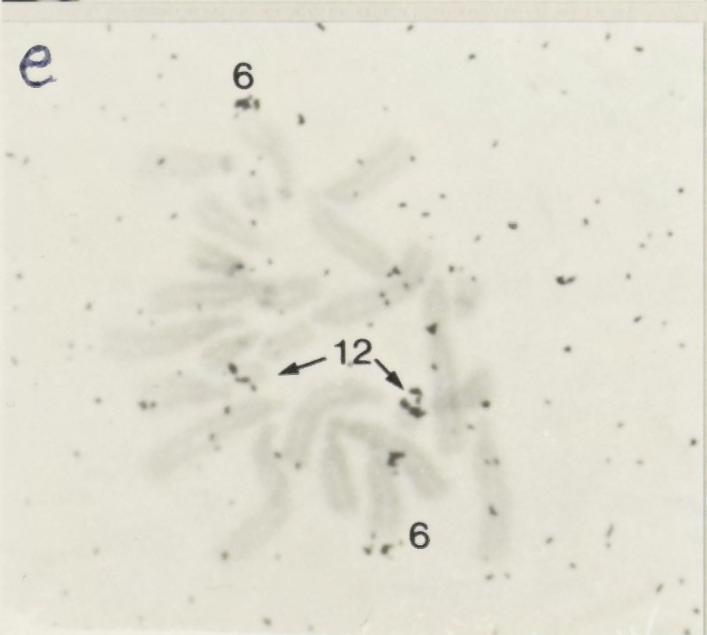
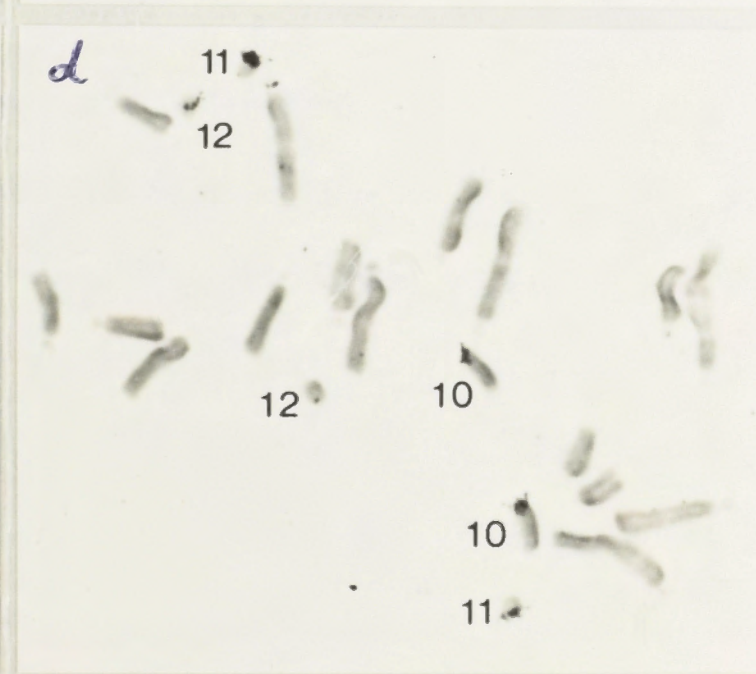
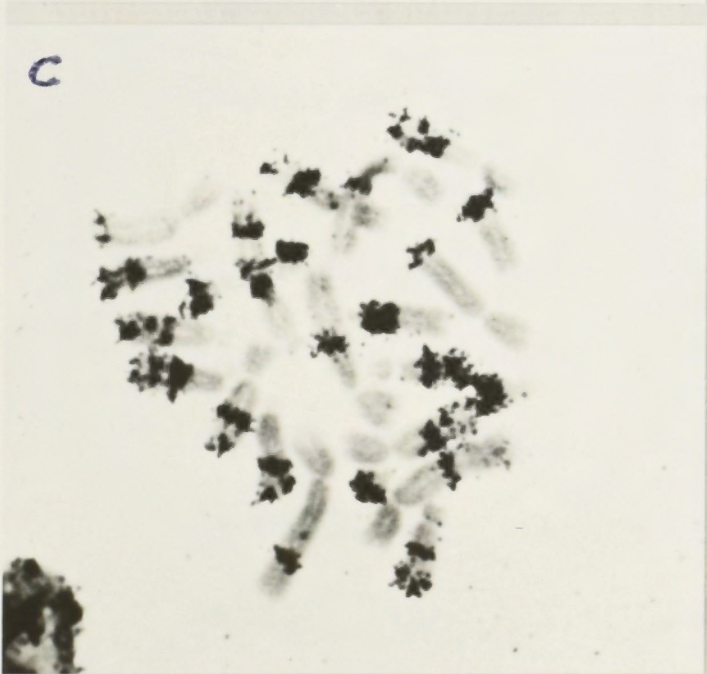
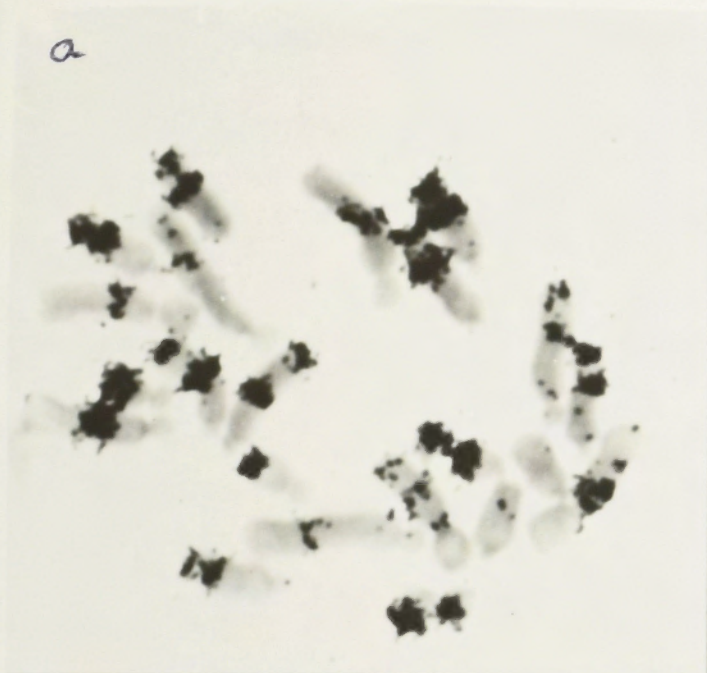




Fig. 2 In situ hybridization utilizing  $^3\text{H}$  c-RNAs from the 168 bp or 144 bp sequences. Top left, S.E.A. individual from the Araluen population hybridized to 168 bp sequence (14 day exposure); top right, S.E.A. individual from the Lakes Entrance population hybridized to 168 bp sequence (22 day exposure); middle left, Moreton individual from the Peregrine population hybridized to 168 bp sequence (5 day exposure); middle right, Southern Torresian individual from the Bongmuller population hybridized to 168 bp sequence (75 day exposure); bottom left, Papuan Torresian individual from the Morehead population hybridized to 168 bp sequence (198 day exposure); bottom right, F1 individual from a Moreton X Daintree cross hybridized to 144 bp sequence (3 day exposure), the numbered chromosomes originate from the Daintree genome.







telomere of chromosome 11 (Table 1). Finally, the 144 bp sequence isolated from the Daintree taxon was shown to be present in the procentric regions of chromosomes 2-7, 9 and 10 (Fig. 2f).

iv) Quantitative measurements for the 168 bp and 144 bp sequences

Table 1 also lists the copy numbers of the 168 bp and 144 bp families for some of the C. captiva populations. The 168 bp family is represented by approximately 150,000, 35,000 and 4,000 copies per haploid genome in the Gundaroo, Bongmuller and Daintree populations, respectively. In addition, the in situ data and preliminary quantitative estimates suggest that the Lakes Entrance and Peregrine individuals have relatively the same number of copies as the Gundaroo sample. Likewise, the Papuan Torresian individuals appear to possess approximately the same amount of the 168 bp repeats as the Bongmuller animals. No copies of this repeat family were detected in C. species nova 1.

The 144 bp Daintree repeat is apparently present in only the Daintree and Papuan Torresian taxa. Individuals from these two taxa have  $2.0 \times 10^6$  and  $1.0 \times 10^5$  copies per haploid genome, respectively (Table 1).

v) Sequence characteristics of the highly repeated DNA sequences

Arnold et al. (1985a, b) have detailed the sequence variation present in the 168 bp repeat which was assayed cytologically in this study. The 144 bp repeat is shown in Figure 3. No restriction endonuclease recognition sites were

Fig. 3 Nucleotide sequence of the 144 bp sequence isolated from the Daintree taxon.



detected within the sequence. The repeat (A-T-G-T) (583) and possesses a region of 1961 symmetry, beginning at position 1 through to position 21.

# DISCUSSION

Evolution of highly repeated DNA and some related aspects within *Callosia*

## 1. 163 bp Sequence Family

The pattern of the cytological distribution and the sequence copy number of the 163 bp repeats, within the various taxa, are correlated to some degree with the amount of heterozygosity of the

GAAGAGAGCAATATTGGTGACGAGAGGACGGGTGGA  
AGTGGTGGAGGAGCCGCCCGTATTCCATAAAGGAGA  
ATTTTAGGTATGTTTAGCGGTATTAACGCTGAACC  
AAGCAATCAAATGCATCTTATTAGTATAATAAATTC

regions of constitutive heterozygosity in these two races. Likewise, the relative disposition of this sequence on the *Torraxia* chromosome corresponds in the data to the presence of terminal blocks of C-banded chromatin. However, for three some of *Parus* *Torraxia* there is either a polymorphism for the presence of a terminal C-band or the 163 bp sequence is present in a region which is cytologically defined as heterozygous. Furthermore, in each of the *Torraxia*, *Parus* and *S.T.S.* populations examined, a proportion of the C-banded positive regions did not hybridize with the 163 bp sequence. In the *Torraxia* genome only a fraction of the constitutive heterozygosity contained this sequence family.

The extensive intertaxonomic variation, in both the amount

detected within the sequence. The repeat is A-T rich (58%) and possesses a region of dyad symmetry, beginning at position 4 through to position 21.

## DISCUSSION

Evolution of highly repeated DNA and taxa relationships within Caledia

### i) 168 bp Sequence Family

The pattern of the cytological distribution and the sequence copy number of the 168 bp repeats, within the various taxa, are correlated to some degree with the amount and distribution of their C-band positive regions. Thus, the numerous interstitial locations of the 168 bp sequence in the S.E.A. and Moreton individuals reflect the presence of equivalent interstitial regions of constitutive heterochromatin in these two races. Likewise, the telomeric disposition of this sequence on the Torresian chromosomes corresponds in the main to the presence of terminal blocks of C-banded chromatin. However, for chromosome 6 of Papuan Torresian there is either a polymorphism for the presence of a telomeric C-band or the 168 bp sequence is present in a region which is cytologically defined as euchromatin. Furthermore, in each of the Torresian, Moreton and S.E.A. populations examined, a proportion of the C-band positive regions did not hybridize with the 168 bp sequence. In the Torresian genome only a fraction of the constitutive heterochromatin contained this sequence family.

The extensive intertaxonomic variation, in both the amount



and the distribution of the 168 bp sequences, indicates that the evolution of this family must have involved amplification/reduction events, as well as the dissemination of the repeats to numerous chromosomal locations. The widespread occurrence of a specific sequence family within a genome its concomitant multiplicity and the marked level of homology between individual repeats from divergent genomic localities are thought to be the result of a process of "concerted evolution" (Zimmer et al., 1980; Arnheim, 1983). One mechanism which may account for a change in the amount of tandemly arrayed sequences is unequal crossover (Smith, 1976). As a result of this process, one chromatid would accumulate a larger segment of these repeats while the other strand would lose a corresponding portion of the tandem array. In addition to the process of amplification/reduction, the 168 bp sequences have been spread to novel chromosomal sites within the Caledia genome. Botchan (1974) proposed that the bovine satellite I DNA had most likely evolved as a consequence of a "rolling circle" mechanism of amplification. This is postulated to involve the production of rings of satellite DNA, which are excision products from reciprocal intrastrand recombination events. According to this model, following amplification the satellite array may reintegrate into the germ line genome. This process would result in the spread of the sequence family if its reintegration were to occur in a nonhomologous chromosomal location. A second mechanism which may account for the interchromosomal movement of these sequences would involve the action of transposable elements. For instance, elements such as the prokaryotic IS



sequences have the ability to transpose entire chromosome segments from one position to another (see Iida et al., 1983 for review). If a single repeat or a tandem set of sequences were positioned between two such elements, the transposition of these elements could include the repeat sequence as well.

The intertaxonomic variation in the amounts of the 168 bp repeats shows an overall negative correlation with the patterns of reproductive isolation and genic divergence detected within this group. Thus, the reproductively isolated Daintree taxon has the least amount of this sequence family in comparison to the Torresian and Moreton taxa. These major taxa groups also show this same pattern in terms of their genic divergence as estimated from allozyme data (Daly et al., 1981). Even so, there are striking examples where the patterns of reproductive isolation, genic divergence and the evolution of the 168 bp DNA family do not correlate. For instance, individuals from the S.E.A. population at Lakes Entrance possess relatively the same amount of the 168 bp sequence as the Moreton individuals. The Lakes Entrance population is also allozymically equivalent to the Moreton taxon (P. Wilkinson, unpub. data). When individuals from the Lakes Entrance population are crossed with Moreton individuals there is a 42% reduction in viability in the F<sub>2</sub> generation (Coates and Shaw, 1984; Shaw et al., in prep.). Most of this reduction can be explained by the structural differences which exist between the chromosomes from these two taxa in terms of centromere position. Likewise, the Torresian and Moreton taxa are distinctive in the disposition and amount of the 168 bp



family within their respective genomes. When these two taxa are crossed there is a 50% reduction in the viability of the backcross progeny and a 100% reduction in the F2 generation (Shaw and Wilkinson, 1980). When the cause of this hybrid inviability was analyzed, it was concluded that the causal factors pertained to the genic and structural differences which are evident between the two taxa (Shaw et al., 1982; Shaw and Coates, 1983; Coates and Shaw, 1984). However, it is possible that the reduction in viability, in the Moreton X Torresian crosses, attributed to the "genic" component could be explained in part by the differences in the amount and/or distribution of the 168 bp sequences. At the present time it is not possible to separate the effects of these two components on hybrid viability.

ii) 144 bp Sequence Family

The 144 bp repeat family was detected in only the Daintree and Papuan Torresian genomes. There are two possible explanations which can account for the occurrence of this sequence in these two groups. First, the present day distribution could reflect the restriction of this repeat family to the lineage which initially gave rise to the Daintree and Papuan Torresian types. However, there are good grounds for concluding that the lineage which gave rise to the Papuan Torresian taxon was also responsible for the formation of the Southern and Northern Torresian forms since there are extremely low levels of allozymic divergence between these three taxa (Table 2; P. Wilkinson, unpub. data). If this is indeed the case, whilst the 144 bp sequence was amplified (or maintained) in

Table 2. Allozymic divergence between Caledia taxa (P.  
Wilkinson, unpub. data) expressed as Nei (Nei, 1972) genetic  
distance ( $\bar{D}$ ) values.



TAXON	(1)	(2)	(3)	(4)	(5)	(6)
<u>C. species nova</u> 1	(1)	-				
Papuan Torresian	(2)	0.450	-			
Northern Torresian	(3)	0.630	0.078	-		
Southern Torresian	(4)	0.569	0.045	0.035	-	
Moreton	(5)	0.648	0.190	0.208	0.157	-
Daintree	(6)	0.446	0.185	0.369	0.273	0.343

the Papuan Torresian individuals, it must have been lost from the other two Torresian types.

A second interpretation of the distribution of this sequence is that it is a result of introgression from the Daintree genome into the genome of the progenitor of the Papuan Torresian taxon. Individuals of this taxon certainly show a much lower level of allozymic divergence from the Daintree type in comparison with the other two Torresian taxa (Table 2; P. Wilkinson, unpub. data). Furthermore, the Nei  $\bar{D}$  value (0.185) for the Daintree vs Papuan Torresian taxa is equivalent to the divergence between the latter taxon and the Moreton population ( $\bar{D}=0.19$ ) and the Daintree taxon is known to be a sibling species of the Torresian and Moreton taxa (Shaw et al., 1980). It should be possible to distinguish between these two hypotheses through the use of additional highly repeated probes.

Various authors have suggested that repeated DNA families may be selfish (Doolittle and Sapienza, 1980) or parasitic (Orgel and Crick, 1980) in their patterns of replication, neutral in their phenotypic expression (Doolittle and Sapienza, 1980; Orgel and Crick, 1980; Miklos and Gill, 1981) or the results of a process which amplifies "non-adaptive" variants and distributes them between different chromosomes (Dover, 1982; Ohta and Dover, 1984). These hypotheses regard highly repeated sequences to be functionless although it has been suggested that the accumulation of non-adaptive variants, by the process of "molecular drive" (Dover, 1982), might give rise to reproductive isolation (Ohta and Dover, 1984). In the following discussion we consider some of the effects which may be attributed to the quantitative or



qualitative characteristics of highly repeated DNA sequences.

#### Potential Effects of Tandemly Repeated DNA Families

##### i) Effects due to quantitative flux

The late replication of heterochromatin has been well documented in a number of organisms (Lima-De-Faria, 1959, Lima-De-Faria and Jaworska, 1968). This differential pattern of replication involves not only entire heterochromatinized chromosomes (eg. sex chromosomes), but also heterochromatic versus euchromatic segments on the same chromosome (Lima-De-Faria, 1959). By examining specific satellite DNA fractions, Bostock et al. (1976) demonstrated that all the various components were synthesized during S-phase. However, the different fractions showed distinct and rapid changes in their rates of synthesis at specific times during this phase. In Caledonia, we have demonstrated that at least two highly repeated sequence families are associated with some of the constitutive heterochromatin from different taxa. If these and other sequences are differentially replicated, then one can envisage that the large scale amplification of a repeat family might result in a change in the cell cycle time. Indeed, Nagl (1974) has reported a negative correlation between the amount of heterochromatin within plant genomes and the duration of the cell cycle. This finding suggests that the amplification/reduction of a specific sequence family has the potential for modulating the cell cycle parameters. In the present case, such a change could result in a differential pattern of development and thereby affect a change



in the phenotype of the grasshopper. Although this is speculative, a finding by Shaw et al. (1985) suggests a potential for this sort of an effect. In a comparison of embryonic weight, it was discovered that Moreton hatchlings were 20% lighter than the Torresian offspring. This result was not due to any difference in the overall time of embryogenesis. One might predict such a difference if the presence of large blocks of heterochromatin, such as are found in the Moreton individuals, causes a fluctuation in the cell cycle time. A detailed examination of this cellular parameter is currently underway.

One effect which is commonly correlated with the presence of large blocks of heterochromatin is the redistribution of chiasmata (Miklos and Nankivell, 1976). This effect is most often associated with polymorphic rather than fixed systems (John and King, 1985). Miklos and Nankivell (1976) showed that the telomeric and centromeric blocks of heterochromatin in the grasshopper species Atractomorpha similis possessed numerous copies of a cryptic satellite sequence. Their findings concerning the pattern of chiasma formation in A. similis in comparison to the other species of this genus (A. australis, A. species-1) led them to conclude that "... the three species are utilizing repeated DNA as a strategy to alter their frequencies and positions of chiasmata, and thus to alter one of the major sources determining variability in the next generation." Detailed analyses of chiasma distribution have been carried out for several of the Caledia taxa, including the Daintree and Moreton types (Shaw and Knowles, 1976; Coates and Shaw, 1982). However, for all of the chromosomes examined, no effect due to



the presence of the heterochromatic blocks was detected. We therefore conclude that the highly repeated sequences which form a part of these blocks are not acting to alter the pattern of recombination to any significant level.

ii) Effects due to nucleotide sequence

It is well established that the chromosomes are attached to certain areas of the inner nuclear envelope (see Franke, 1974 for review). Furthermore, it has been demonstrated in the mouse that at least a portion of the chromatin associated with the nuclear envelope is made up of highly repeated DNA sequences (Rae and Franke, 1972). These and other findings concerning this association led Franke (1974) to ask: "Does the nuclear membrane contain compounds, probably proteins, that specifically complex with certain chromosomal regions and exclude others?" This sort of specific affiliation has indeed been suggested by studies of in vitro binding of nucleic acids to the nuclear membrane (Kubinski et al., 1972). With this in view it is worth noting that Arnold et al. (1985a) detected a conserved 18 bp sequence which was part of a region of dyad symmetry within the 168 bp repeat. This 18 bp sequence was conserved within individuals and populations as well as between different races and subspecies. It is plausible that the presence of the two-fold symmetry, characteristic of DNA-protein binding sites (Siebenlist et al., 1980), and the conservation of a portion of this region is indicative of its functioning in complexing with the nuclear envelope. An alternative putative effect is that this region may play a role in effecting gene expression. Appels and Peacock



(1978) have pointed out that "... in cases in which a repeating sequence resembles the binding site of a particular protein, the functions of this protein will be markedly affected." A particularly lucid example of such an effect has been documented in vitro for the Xenopus satellite I DNA sequences (Andrews et al., 1984). In both HeLa S100 and oocyte microinjection transcription assays, the satellite sequences inactivated 5S r-RNA transcription. This effect was apparent whether the satellite sequences were in a cis or trans relationship with the ribosomal genes. One possible explanation for this competitive inactivation was the presence of an RNA polymerase III transcription unit in the satellite repeats which was 10-25-fold more efficient than the r-RNA gene in the competition experiments (Andrews et al., 1984).

Although a specific nuclear interaction has yet to be demonstrated for the C. captiva repeated DNA elements, the potential functional importance of such interactions merits further study.

Acknowledgements. We wish to express our gratitude to P. Wilkinson for allowing us to use her unpublished allozyme data. F. Arnold typed numerous drafts of the manuscript. We thank CSIRO, Division of Plant Industry and in particular R. Appels for making laboratory facilities available. M.A. was supported by an Australian National University Scholarship. R. Appels, P. Wilkinson and L. Moran provided invaluable technical assistance. We also thank B. John for reviewing the manuscript. G. Brown and



M. Whitaker aided in figure preparation and photography, respectively.

#### REFERENCES

- Andrews DL, Millstein L, Hamkalo BA, Gottesfeld JM (1984)  
Competition between *Xenopus* satellite I sequences and Pol III  
genes for stable transcription complex formation. NAR 12:7753-  
7769
- Appels R (1982) The molecular cytology of wheat-rye hybrids.  
Int Rev Cytol 80:93-132
- Appels R, Driscoll C, Peacock WJ (1978) Heterochromatin and  
highly repeated DNA sequences in rye (*Secale cereale*).  
Chromosoma 70:67-89
- Appels R, Peacock WJ (1978) The arrangement and evolution of  
highly repeated (satellite) sequences with special reference to  
*Drosophila*. Int Rev Cytol Suppl 8:69-126
- Arnheim N (1983) Concerted evolution of multigene families. In:  
Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer  
Associates, Inc., Sunderland, p 38-61
- Arnold ML, Appels R, Shaw DD (1985a) The heterochromatin of  
grasshoppers from the *Caledia captiva* species complex I.  
Sequence evolution and conservation in a highly repeated DNA  
family. Mol Biol Evol, in press
- Arnold ML, Appels R, Shaw DD (1985b) Evolution and conservation  
in a highly repeated DNA family from the grasshopper *Caledia*  
*captiva*. Cytobios, in press



- Bishop JO, Robertson FW, Burns JA, Melli M (1969) Methods for the analysis of deoxyribonucleic acid - ribonucleic acid hybridization data. *Biochem J* 115:361-370
- Bostock CJ, Christie S, Lauder IJ (1976) S-phase patterns of replication of different satellite DNAs in three species of *Dipodomys* (Kangaroo Rat). *J Mol Biol* 108:417-433
- Botchan MR (1974) Bovine satellite I DNA consists of repetitive units 1400 base pairs in length. *Nature* 251:288-292
- Coates DJ, Shaw DD (1982) The chromosomal component of reproductive isolation in the grasshopper *Caledia captiva* I. Meiotic analysis of chiasma distribution patterns in two chromosomal taxa and their F1 hybrids. *Chromosoma* 86:509-531
- Coates DJ, Shaw DD (1984) The chromosomal component of reproductive isolation in the grasshopper *Caledia captiva* III. Chiasma distribution patterns in a new chromosomal taxon. *Heredity* 53:85-100
- Cohen EH, Bowman SC (1979) Detection and location of three simple sequence DNAs in polytene chromosomes from virilis group species of *Drosophila*. *Chromosoma* 73:327-355
- Daly JC, Wilkinson P, Shaw DD (1981) Reproductive isolation in relation to allozymic and chromosomal differentiation in the grasshopper *Caledia captiva*. *Evolution* 35:1164-1179
- Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* 284:601-603
- Dover G (1982) Molecular drive: a cohesive mode of species evolution. *Nature* 299:111-116
- Franke WW (1974) Structure, biochemistry, and functions of the nuclear envelope. *Int Rev Cytol Suppl* 4:71-236



- Gall JG, Cohen EH, Atherton DD (1974) The satellite DNAs of *Drosophila virilis*. Cold Spring Harb Symp Quant Biol 38:417-421
- Iida S, Meyer J, Arber W (1983) Prokaryotic IS elements. In: Shapiro JA (ed) Mobile genetic elements. Academic Press, New York, p 159-221
- John B, King M (1985) The inter-relationship between heterochromatin distribution and chiasma distribution. Genetica, in press
- Kubinski H, Gibbs P, Kasper CB (1972) Interactions of nucleic acids and cellular membranes: behavior of membranes in CsCl density gradients. Biochem Biophys Acta 281:244-252
- Kurnit DM, Shafit BR, Maio JJ (1973) Multiple satellite deoxyribonucleic acids in the calf and their relation to the sex chromosomes. J Mol Biol 81:273-284
- Laskey RA (1980) The use of intensifying screens or organic scintillators for visualizing radioactive molecules resolved by gel electrophoresis. Meth Enzymol 65:363-371
- Lima-De-Faria A (1959) Differential uptake of tritiated thymidine into hetero- and euchromatin in *Melanoplus* and *Secale*. J Biophys Biochem Cytol 6:457-466
- Lima-De-Faria A, Jaworska H (1968) Late DNA synthesis in heterochromatin. Nature 217:138-142
- Messing J, Vierira J (1982) A new pair of M13 vectors for selecting either strand of double-digest restriction fragments. Gene 19:269-276

- Miklos GLG, Gill AC (1981) The DNA sequences of cloned complex satellite DNAs from hawaiian *Drosophila* and their bearing on satellite DNA sequence conservation. *Chromosoma* 82:409-427
- Miklos GLG, Nankivell RN (1976) Telomeric satellite DNA functions in regulating recombination. *Chromosoma* 56:143-167
- Nagl W (1974) Role of heterochromatin in the control of cell cycle duration. *Nature* 249:53-54
- Nei M (1972) Genetic distance between populations. *Amer Natur* 106:283-292
- Novak U (1984) Structure and properties of a highly repetitive DNA sequence in sheep. *NAR* 12:2343-2350
- Ohta T, Dover GA (1984) The cohesive population genetics of molecular drive. *Genetics* 108:502-521
- Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. *Nature* 284:604-607
- Pardue ML, Gall JG (1970) Chromosomal localization of mouse satellite DNA. *Science* 168:1356-1358
- Peacock WJ, Appels R, Dunsmuir P, Lohe AR, Gerlach WL (1977) Highly repeated DNA sequences: Chromosomal localization and evolutionary conservatism. In: Brinkley BR, Porter KR (eds) *International cell biology*. Rockefeller University Press, New York, p 494-506
- Peacock WJ, Brutlag D, Goldring E, Appels R, Hinton CW, Lindsley DL (1974) The organization of highly repeated DNA sequences in *Drosophila melanogaster* chromosomes. *Cold Spring Harb Symp Quant Biol* 38:405-421



- Peacock WJ, Lohe AR, Gerlach WL, Dunsmuir P, Dennis ES, Appels R  
(1978) Fine structure and evolution of DNA in heterochromatin.  
Cold Spring Harb Symp Quant Biol 42:1121-1135
- Rae PMM, Franke WW (1972) The interphase distribution of  
satellite DNA-containing heterochromatin in mouse nuclei.  
Chromosoma 39:443-456
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-  
terminating inhibitors. PNAS 74:5463-5467
- Shaw DD (1976) Population cytogenetics of the genus *Caledia*  
(Orthoptera: Acridinae) I. Inter- and intraspecific karyotype  
diversity. Chromosoma 54:221-243
- Shaw DD, Coates DJ (1983) Chromosomal variation and the concept  
of the coadapted genome - a direct cytological assessment. In:  
Brandham PE, Bennett MD (eds) Kew Chromosome conference II. G  
Allen and Unwin, London, p 207-216
- Shaw DD, Coates DJ, Arnold ML, Wilkinson P (1985) Temporal  
variation in the chromosomal structure of a hybrid zone and its  
relationship to karyotypic repatterning. Heredity, in press
- Shaw DD, Knowles GR (1976) Comparative chiasma analysis using a  
computerised optical digitiser. Chromosoma 59:103-127
- Shaw DD, Moran C, Wilkinson P (1980) Chromosomal reorganization,  
geographic differentiation and the mechanism of speciation in  
the genus *Caledia*. In: Blackman RL, Hewitt GM, Ashburner M  
(eds) Insect cytogenetics, Symp Roy Entomol Soc London, No 10.  
Blackwells, Oxford, p 171-194
- Shaw DD, Webb GC, Wilkinson P (1976) Population cytogenetics of  
the genus *Caledia* (Orthoptera: Acridinae) II. Variation in the  
pattern of C-banding. Chromosoma 56:169-190



- Shaw DD, Wilkinson P (1980) Chromosome differentiation, hybrid breakdown and the maintenance of a narrow hybrid zone in *Caledia*. *Chromosoma* 80:1-31
- Shaw DD, Wilkinson P, Coates DJ (1982) The chromosomal component of reproductive isolation in the grasshopper *Caledia captiva* II. The relative viabilities of recombinant and non-recombinant chromosomes during embryogenesis. *Chromosoma* 86:533-549
- Siebenlist U, Simpson RB, Gilbert W (1980) *E. coli* RNA polymerase interacts homologously with two different promoters. *Cell* 20:269-281
- Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* 191:528-535
- Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98:503-517
- Strachan T, Webb D, Dover G (1985) Transition stages during sequence turnover in multiple-copy DNA families in *Drosophila*. *Cell*, in press
- Sutton WD, McCallum M (1972) Related satellite DNA's in the genus *Mus*. *J Mol Biol* 71:633-656
- Trick M, Dover GA (1984) Unexpectedly slow homogenisation within a repetitive DNA family shared between two subspecies of tsetse fly. *J Mol Evol* 20:322-329



Webb GC (1976) Chromosome organization in the Australian plague locust, *Chortoicetes terminifera*. I. Banding relationships of the normal and supernumerary chromosomes. *Chromosoma* 55:229-246

Zimmer EA, Martin SL, Beverley SM, Kan YW, Wilson AC (1980) Rapid duplication and loss of genes coding for the  $\alpha$  chains of hemoglobin. *PNAS* 77:2158-2162



## CHAPTER 5. Bam Family Repeated Sequences - Repeats that are Related to the Taq Family Sequences

### 5.1 Introduction

The major analysis in this thesis was carried out on the 168 bp Taq family. In a protocol designed to recover tandem arrays of the 168 bp repeat, a related sequence family was discovered. The new family of sequences is defined by the restriction endonuclease Bam HI and has a degree of homology to the 168 bp family reflected in their ability to cross-hybridize. However, the degree of homology at the nucleotide sequence level is too low to consider the Bam sequences to be a part of the Taq family. Thus, the Bam family of sequences is discussed separately in this chapter.

### 5.2 Results and Discussion

The digestion of C. captiva DNA with the restriction endonuclease Bam HI resolves a ladder of bands after hybridization with a 168 bp family probe (Fig. 5-1). Initially, this ladder of bands was assumed to represent sequence variants within the 168 bp family. As can be seen in Figure 5-1, the Daintree and Torresian samples are mainly in tandem arrays which appear as defined bands following Bam HI restriction. In contrast, the Moreton and, in particular, the S.E.A. samples demonstrate less distinct bands (Fig. 5-1) and a high background of hybridization to heterogeneous DNA sequences. This would suggest the presence of homologous sequences throughout a large range of size classes, which is indicative of the sequences being dispersed as well as clustered. These repeats were cloned by



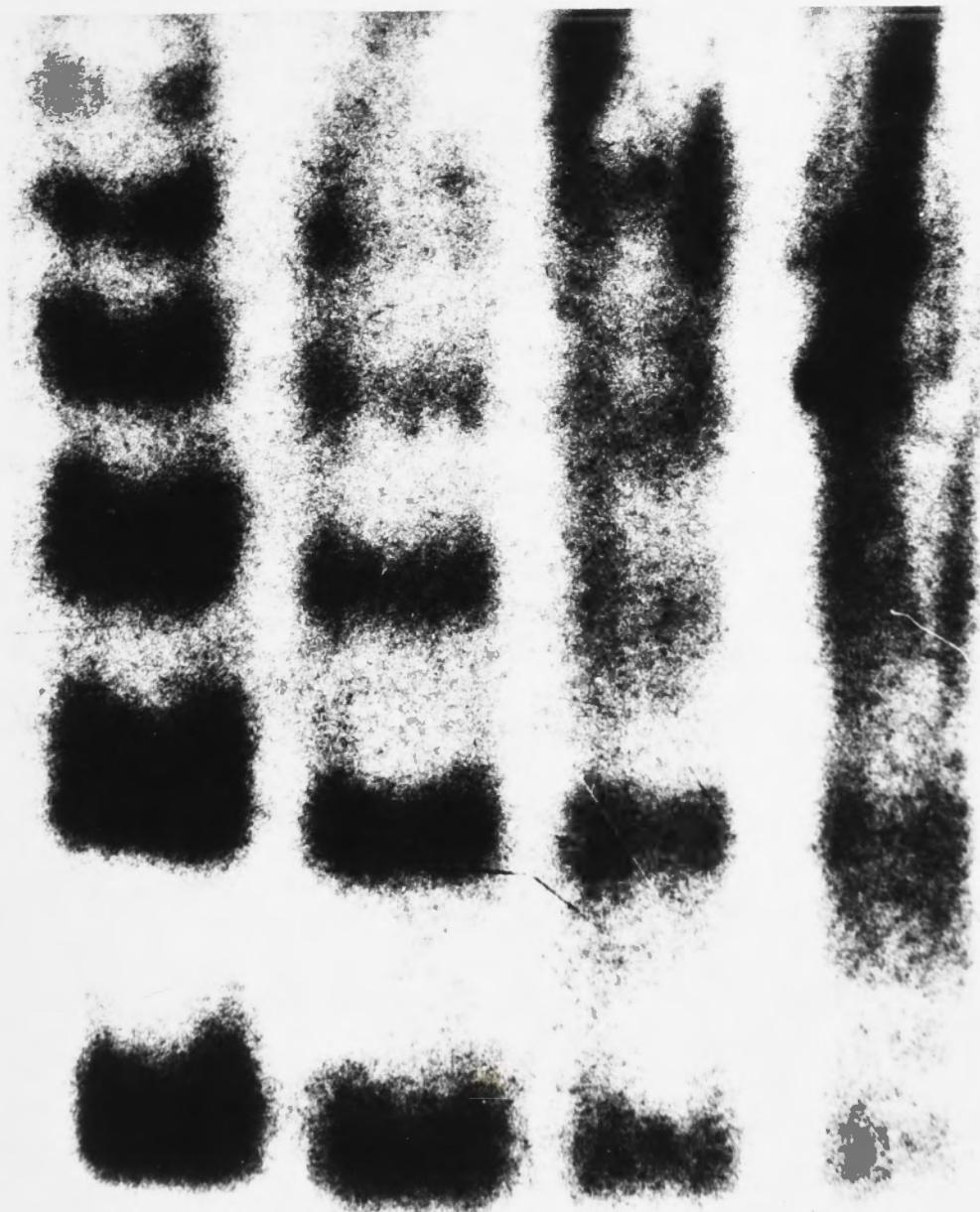
Figure 5-1. Bam HI restricted DNA hybridized to a probe from the 168 bp sequence. a. Daintree sample, b. Southern Torresian sample, c. Moreton sample, d. S.E.A. sample. Lowest bands in each lane are approximately 180 bp in length. Upper portion of X-ray, which contained extensive hybridization, was deleted.

a

b

c

d





ligating the Bam HI digested DNA directly into an M13 vector and then selecting the bacteriophage which showed homology with the 168 bp probe (see Chapter 2). Eight recombinant bacteriophage were isolated and used for the dideoxy sequencing reactions. The sequence data from the bacteriophage are given in Figures 5-2 and 5-3.

Representatives of the Bam family of sequences were recovered from the South-east Australian (S.E.A. - Gundaroo population), Moreton (Peregian Beach population), Southern Torresian (Bongmuller and Insulator Creek populations), Papuan Torresian (Morehead population) and Daintree (Daintree population) taxa. Individual repeats were isolated (Fig. 5-2) except in the case of the Moreton sample where a 774 bp sequence was examined (Fig. 5-3). This sequence consisted of four complete monomer units along with 42 bp from a fifth repeat. The consensus sequence derived from the seven repeats of the S.E.A., Torresian and Daintree taxa and the first entire repeat (Fig. 5-3; beginning at position 43) from the Moreton tandem array is shown in Figure 5-2. This consensus sequence is 183 bp in length, 54% A+T and has 75 variable positions. Similarly, the consensus sequence for the monomer units from the Moreton clone is 185 bp long and is also 54% A+T (Fig. 5-3). However, this sequence has a total of only 29 variable positions. When the variable sites in the tandem array are added to the sites identified in Figure 5-2, a total of 84 variable positions can be identified. Each of the repeat units illustrated in Figure 5-2 were examined in pairwise comparisons, as were the repeats from the Moreton tandem array, and the results are given in Table 5-1.

Figure 5-2. Nucleotide sequence for the Bam family repeats from the C. captiva taxa. The bold-faced sequence is the consensus for the repeats. 1. Gundaroo population (S.E.A.), 2. Peregrine Beach population (Moreton), 3. Bongmuller population (Southern Torresian), 4. Insulator Creek population (Southern Torresian), 5. Morehead population (Papuan Torresian) and 6-8. Daintree population (Daintree). Asterisks (\*) indicate the variable positions. Dashes (-) denote that the position lacks a base. Lower case letters indicate that more than one choice is apparent for the consensus sequence at the given position.



1	10	20	30	40	50	60	70	80	90
:	:	:	:	:	:	:	:	:	:
		*	**	**	*	***	**	*	**
				*	*	*	*	*	*

TCCTTTCTCCTAATCGGCTCGGTAATCTGCCTTCTTAATTTGTTACGTAGCATTT-GGATGTGGGTGCTATACCTATACAAAGTTTCATAAGGCGCACT  
 1 TCCTTTCTCCTAATCGGCTCTGTAATCTGCCTTCTTAATTTGTTACGTAGCATTT-GGATGTGGGTGCTAAACCTATACAAAGTTTCATAAGGCG-ACT  
 2 TCCTTTCTCCTAATCGGCTCTGTAATCTGCCTTCTTAATTTGTTACGTAGCATTT-GGATGTGGGTGCTATACATATACAAAGTTTCATAAGGCGCACA  
 3 TCCTTTCTCCTAATCGGCTCGGTAATCTACCTTCTTAATTTGTTAGGTAGCATTTTGGATGTGGGTGCTATAC-TGTACAAAATTTTCATTAGGCGCACT  
 4 TCCTTTCTCCTAATCGGCTCGGTAATCTGCCTTCTTAATTTGTTACGTAGCATTT-CAACGAGGGTGCTATACCTATACAAAGTTTCATAAGGCGCACT  
 5 TCCTTTCTCCTAATCGGCTCGGTAATCTGCCTTCTTAATTTGTTACGTAGCATTT-GGATGTGGGTGCTATATCTATACAAAGTTTCATTAGGCGCACT  
 6 TCCTTTCTCCTAATCGGCTCGGTAATCTGCTTTCTTAATTTGTTCCGTAGCATTT-GGATGTGGGTGCTACACCTATACAAAGTTTCATAAGGGGCACA  
 7 TCCTTTCTCCTAATCGGCTCGGTAATCTGCCTTCTTAAGTTTTTGGGTAGCATTT-GGATGTGGGTGCTATACCTATACAAAGTTTCATAAGGGGCAGC  
 8 TCCTTTCTCCTAATCGGCTCGGTAATCTGCCTTCTTCATTTGTTACGTAGCAATT-GGAAGTGTGTGCTATACCTAAACAAAGTTTCATAAGGCGCAGG

100	110	120	130	140	150	160	170	180	190
:	:	:	:	:	:	:	:	:	:
**	* ***	***** *		***	*****	* ** *	*****	* ** *	** * *
									***** ** *

CTGACAGGACTGTAGTTACAC-GCCCTTGAATATA--GCAGGta-GG-T-CC-TGCCGCGCGAC-GCCGGAAA-TTTTATCGGT-TTTC---GAAGGA  
 1 CTGACAGGACTGTTGTTACAT-GGCCTTGAATATA--AC-----GG-T-CC-TGCCGCCCCGAC-GCAGGAAA-TTTTATTGGT-TT----AGAAGGA  
 2 CTGACGGGACTTTAGTTACTC-ACC-TTGAATATA--GCAGGTAAGGGT-CC-TGCCGG--GAC-GCCGGAAA-TTTTATCGGTGTT-C---GAAGCA  
 3 CTGACAGGACTGTAGTTACAC-GGCCTTGAATATA--TCAGGGA-GG-TTCC-TGCCG-GTGAC-GC-GGAAA-TTTTATCGGT-TT-C---GA-GGA  
 4 CTGACAGGACTGTAGTTACAC-GCCCTTGAATATA--GCAGGGAAGG-T-CCGTGCCACGTGAC-GCTGGAAA-TTTTATCGGT-TTTCGGAG-AGGA  
 5 CTGACAGGACTATAGTTACCC-GCCCTTGAATATA--AC-----GG-T-CC-TGCCGCCCCGAC-GCCGGAAA-TTTTATCGGT-TTTC---GAAGGA  
 6 CTGACGGGACTGTAGTTACAC-GCCCTTGAATATA--AC-----GG-T-CC-TGCCGCGCGAC-GCCGGAAATTTTATCGCT-TTTC---GAAGGA  
 7 CTGACGGGACTGTAAATACTCGG-CCTTGAATATAAAGCAGGT--GG-T-AC-CGCTGCGCGAC-GC--GGCATTTTTACCGCT-TTTC---GAAGGA  
 8 CTGACAAGACTGTAGTTACAC-GCCCTTGAATATAAAGCAGGTA-GG-T-ACGTGCTGCGCGACGGC--GAAATTTTATCGGT-TTTC---GAAGGA

Figure 5-3. Nucleotide sequence for the Bam family tandem array from the Moreton taxon. Note that repeat unit 5 has a Bam HI recognition site beginning at position 40 and, therefore, only the first 42 bases were cloned. The homologous regions to the repeats shown in Figure 5-2 begin at position 43 of the repeats in this figure. The designations are the same as used in Figure 5-2. The two "halves" of the Bam recognition sites that defined this tandem array are at the beginning of repeat 1 (TCC) and the end of repeat 5 (GGA).



	1	10	20	30	40	50	60	70	80	90
	:	:	:	:	:	:	:	:	:	:
		*	*	*	*	*	*	*	*	*
	TCCTGCCGCGCGACGCCGGAATTTTATCGGTTTTTCGAAGCATCCTTTCTCCTAATCGGCTCTGTAATCTGCCTTCTTAATTTGTTACGTAGC									
1	TCCTGCTGCGCGAAGCCG-AAATTTTATCGGTGTTTCGAAGCATCCTTTCTCCTAATCGGCTCTGTAATCTGCCTTCTTAATTTGTTACGTAGC									
2	TCCTGCCGCGCGACGCCGGAATTTTATCGGTGTTTCGAAGCATCCTTTCTCCTAATCGGCTCTGTAATCTGCCTTCTTAATTTGTTACATAGC									
3	TCCTGCCGCGCGACGCCGGAATTTTATCGGTTTTTCGAAGCATCCTTTCTCCTACTCGGCTCGGTAATCTGCCTTCTTAATTTGTTACGTAGC									
4	TCCTGCCGTGCGACGCCGGAATTTTATCGGTTTTTCGAAGCATCCTTTCTCCTAATTGGCTCGGTAATCTGCCTTCTTAATTTGTTACGTAGC									
5	TCCTGCCGCGCGACGCCGGAATTTTATCGGTTTTTCGAAGGA									

	100	110	120	130	140	150	160	170	180
	:	:	:	:	:	:	:	:	:
			**	**	*	*	*	*	***** *
	ATTTGGATGTGGGTGCTATACCTATACAAAGTTTCATAAGGCGCACACTGACGGGACTGTAGTTACaCaCCCTTGAATATAGCAGGTAAGGG								
1	ATTTGGATGTGGGTGCTATACATATACAAAGTTTCATAAGGCGCACACTGACGGGACTTTAGTTACTCACCCTTGAATATAGCAGGTAAGGG								
2	ATTTGGATGTGGGTGCTATAGGTATACAAAGTTTCATATTGCGCACACTGACGGAAGTGTAGTTACTCGCCCTTGAATATAGCAGGTAAGGG								
3	ATTTGGATGTGGGTGCTATACCTATACAAAGTTTCATAAGGCGCACAATGACGGGACTGTAGTTACATGCCCTTGAATATA-----ACGG								
4	ATTTGGATGTGGGTGCTATACCTATACAAAGTTTCATACGGCGCACACTGACGGGACTGTAGTTACACACCCTTGAATATAGCAGGTAAGGG								

Table 5-1. Variation in base pair sequence among the Bam family  
cloned repeats. Numbers represent base pair divergence.



A comparison of the consensus

	<u>RANGE</u>	<u>X</u>
Intratandem repeat <sup>1</sup>	12-22	16(9%)
Intrapopulation <sup>2</sup>	26-28	27(15%)
Intraracial <sup>3</sup>	-	25(14%)
Interracial <sup>4</sup>	23-36	28(15%)
Intersubspecific <sup>5</sup>	15-37	27(15%)
Interspecific <sup>6</sup>	13-40	29(16%)

<sup>1</sup>Peregrine Beach tandem array;

<sup>2</sup>Daintree repeats;

<sup>3</sup>Bongmuller compared to Insulator Creek;

<sup>4</sup>Gundaroo compared to Peregrine Beach, Bongmuller and Insulator Creek compared to Papuan Torresian;

<sup>5</sup>Gundaroo and Peregrine Beach compared to Insulator Creek and Papuan Torresian;

<sup>6</sup>Daintree compared to all other populations.

The ranges and mean number of base pair differences for most of the comparisons are quite similar, however, the "intratandem repeat" values are the lowest.

A comparison of the consensus sequences from the Taq (Chapter 3) and Bam families has shown that the longest contiguous region of homology is only eight base pairs. In addition, the 18 bp conserved region in the Taq family was compared along the entire length of the Bam family consensus sequence. The region that demonstrates the highest level of similarity to the 18 bp Taq sequence is illustrated in Figure 5-4. These two sequences match at 11 of the 18 bp (ie., 61% similarity).

The Bam family sequences from the Daintree taxon were previously placed within the Taq family by Arnold et al. (1985a; see Chapter 3). Thus, they considered these sequences to be highly divergent forms from this family. However, with the isolation of the Bam family sequences from each of the other C. captiva taxa, it has become apparent that the Bam repeats form a related, but highly divergent, group of sequences relative to the Taq family. Therefore, it is expected that these sequences have undergone different amplification/reduction and dissemination events since they diverged from a common ancestral sequence.

Preliminary data on the quantities of the Bam family sequence in each of the C. captiva taxa have suggested that these taxa have similar amounts of the 183 bp repeat. In contrast, the 168 bp Taq family repeats have evidently undergone large fluctuations in both amount and cytological distribution during



Figure 5-4. Comparison of the conserved Taq family 18 bp region (upper sequence) with the 18 bp stretch from the Bam family consensus sequence (lower sequence). Asterisks (\*) indicate positions that match.





the evolution of the C. captiva species complex (Arnold et al., 1985a; see Chapter 3). A number of characteristics of the sequence data for the Taq and Bam repeated DNA families also indicate that the evolutionary histories of these two families have been markedly different. Thus, the divergence between repeats at the level of the individual, population, race and subspecies is approximately 2-4 times greater between the Bam family repeats than for the 168 bp Taq family sequences. Indeed, repeats from the same Bam family tandem array have more interrepeat divergence than do Taq repeats from different subspecies.

One explanation for this difference is that the Bam repeat family might have arisen at an earlier stage in the evolution of the C. captiva complex than the Taq family. The higher levels of divergence in the former family might then be a consequence of the accumulation of random mutations over a relatively longer evolutionary time period. Alternatively, either the entire Taq repeat unit or only a portion of the sequence may be conserved with few or no incorporated mutations. This has in fact been documented in the 168 bp sequence family, for an 18 bp sequence within each of the monomers. The conservation of this region was evident not only within individuals but also between populations, races and subspecies. This pattern of conservation, along with the presence of two-fold symmetry in the region of the 18 bp sequence, has led to the suggestion that this sequence may have functional significance (Chapter 3). Ohta and Dover (1984) have concluded that such a pattern of highly variable and less variable regions within a repeat unit may be the result of gene

conversion. They have stated that, "The domain of conversion can be either shorter or longer than the gene itself. When it is shorter it is possible to evolve highly variable genes that are mosaics of different domains."

Although the Bam repeats have a larger amount of sequence divergence than the Taq sequences, the tandem array from the Moreton taxon possesses the lowest level of sequence divergence relative to the other Bam family comparisons. This observation of less divergence in repeats from the same tandem array in comparison to noncontiguous repeats was also reported by Trick and Dover (1984). They found that there was 24% and 31% divergence between contiguous and noncontiguous repeats, respectively, for a highly repeated DNA family from two subspecies of Tsetse fly. However, as in the Bam family repeats, the variation within the Tsetse fly tandem array overlaps the range of variation found between repeats from different tandem arrays. The lower level of mean variation in these tandem arrays could be the result of a process such as unequal crossing over. Alternatively, the arrays could represent products of a relatively recent amplification event that have not had sufficient time to accumulate mutations to the same level as other repeats within the family. Finally, these results may reflect a greater effect of gene conversion events between contiguous versus noncontiguous repeats. It is of interest to note that the mutation that caused the loss of a Bam HI recognition site resulting in the tandem array is the same in each of the repeats (Fig. 5-3). Thus, at position 41 a cytosine



has been substituted for a guanine. An explanation for the presence of the tetrameric array is that the original mutation at this site produced a dimer of the Bam sequences. Subsequent to this mutation event, the dimer was evidently amplified to form a tetramer. However, one must also account for the finding that position "40" in the tandem array (Fig. 5-3) corresponds to position "1" in the other Bam family sequences (Fig. 5-2). It appears that there has been a further mutation resulting in the production of a Bam HI recognition site approximately 40 bp from the original Bam site in the tandem array.

The amount of sequence divergence between the Daintree sibling species and the other C. captiva taxa overlaps extensively with all of the other pairwise comparisons. This result can be explained on the basis that either the repeats from the Daintree taxon have not had sufficient time to diverge from those of the other C. captiva taxa or that the sequences in each of the taxa are being conserved in terms of their sequence characteristics. Data from studies of Drosophila favor the former possibility. Strachan et al. (1985) have found that two Drosophila highly repeated DNA families possess approximately one order of magnitude greater sequence variation in between species comparisons than is found within a species. This level of divergence was not found when comparisons were made between D. mauritiana and D. simulans and between D. yakuba and D. teissieri. These four taxa are all sibling species within the "melanogaster" species subgroup (Lemeunier and Ashburner, 1976). However, on the basis of polytene chromosome comparisons, the two pairs of species have been placed in different categories within

this subgroup (Lemeunier and Ashburner, 1976). In particular, D. simulans and D. mauritiana have been shown to be homosequential. Therefore, the pattern of divergence within the highly repeated DNA families, detected by Strachan et al. (1985), would appear to be directly correlated with the amount of time since divergence.



## CHAPTER 6: Eco Repeated Sequences

### 6.1 Introduction

In the present chapter, data is presented on three divergent, but apparently related fragments, isolated from the S.E.A. taxon. These sequences are characterized by having Eco RI recognition sites at their 5' and 3' termini and have, therefore, been named "Eco sequences". Although the results from the study of these sequences are preliminary, they are presented because the repeats are interesting in that they appear to be dispersed and to possess an unusual nucleotide structure.

### 6.2 Results and Discussion

Approximately 10-15 kb segments of DNA from C. captiva were cloned into phage  $\lambda$  1059. Three clones were selected on the basis of their homology to a highly repetitive (Cot 0.01) fraction of DNA from the C. captiva genome. One clone ( $\lambda$ SEA 28) was selected for further analysis. Following digestion with Eco RI, three fragments which contained only the grasshopper DNA were isolated. These sequences were designated "a", "b" and "c" and were approximately 500 bp, 900 bp and 5400 bp long, respectively. These three repeats were then cloned into both a plasmid vector and M13 mp10. The plasmid clones pGEA ("a" insert), pGEB ("b" insert) and pGEC ("c" insert) were used to prepare  $^3\text{H}$ -cRNAs. The M13 clones (also called pGEA, pGEB and pGEC in the present discussion) were utilized in sequencing experiments.

Figures 6-1 and 6-2 illustrate the results obtained from in situ experiments with the three sequences. It is apparent that each of the sequences are dispersed throughout the genomes of the

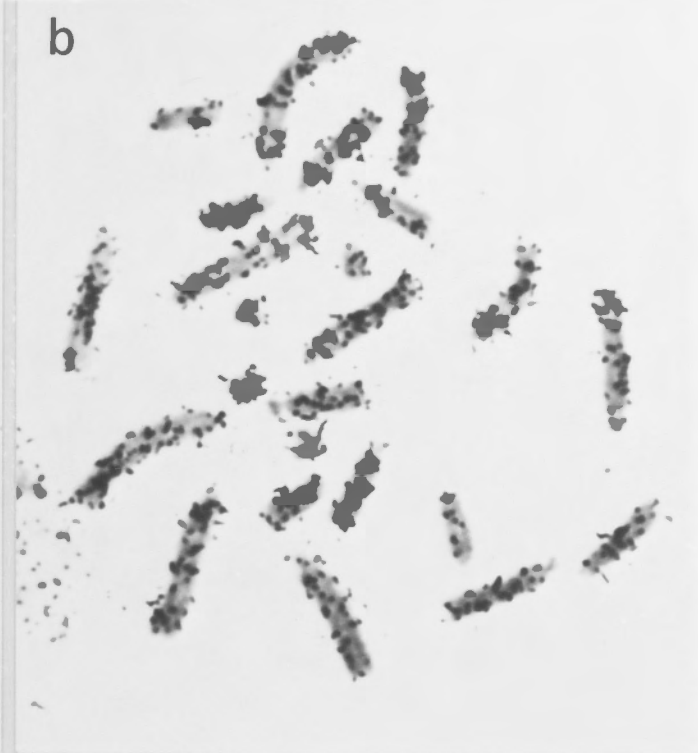
Figure 6-1. In situ hybridizations with the Eco repeats. a.  $^3\text{H}$  c-RNA from pGEA hybridized to an Araluen (S.E.A.) individual (48 day exposure); b.  $^3\text{H}$  c-RNA from pGEC hybridized to a Bongmuller (Torresian) individual (8 day exposure); c.  $^3\text{H}$  c-RNA from pGEB hybridized to an Araluen individual (48) day exposure); d.  $^3\text{H}$  c-RNA from pGEC hybridized to an Araluen individual (7 day exposure).



a



b



c



d

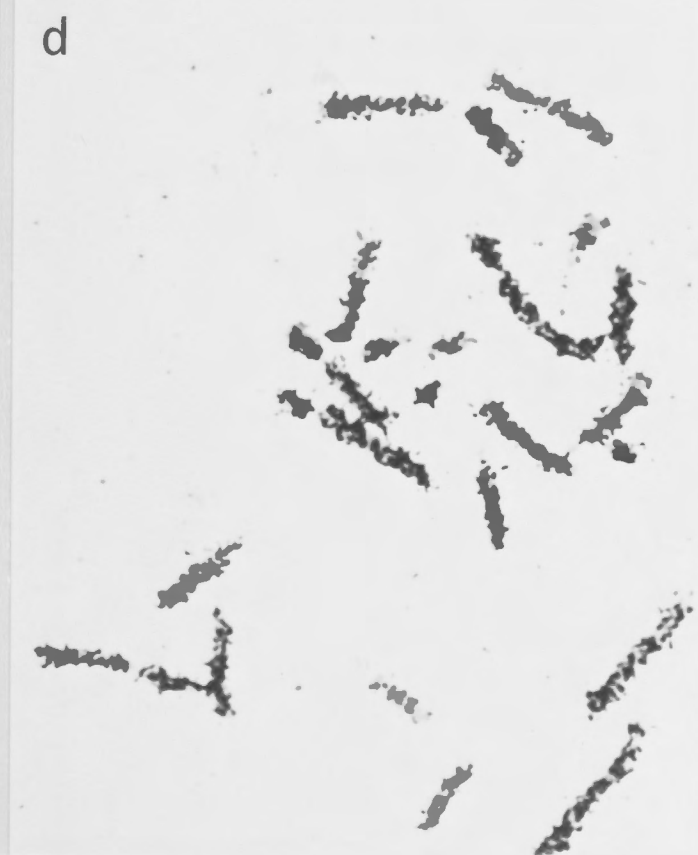
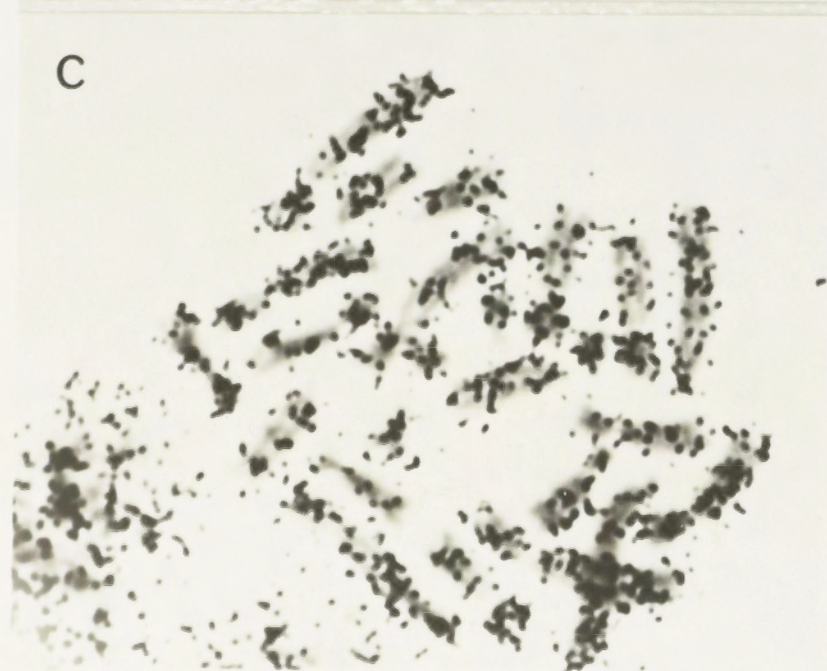
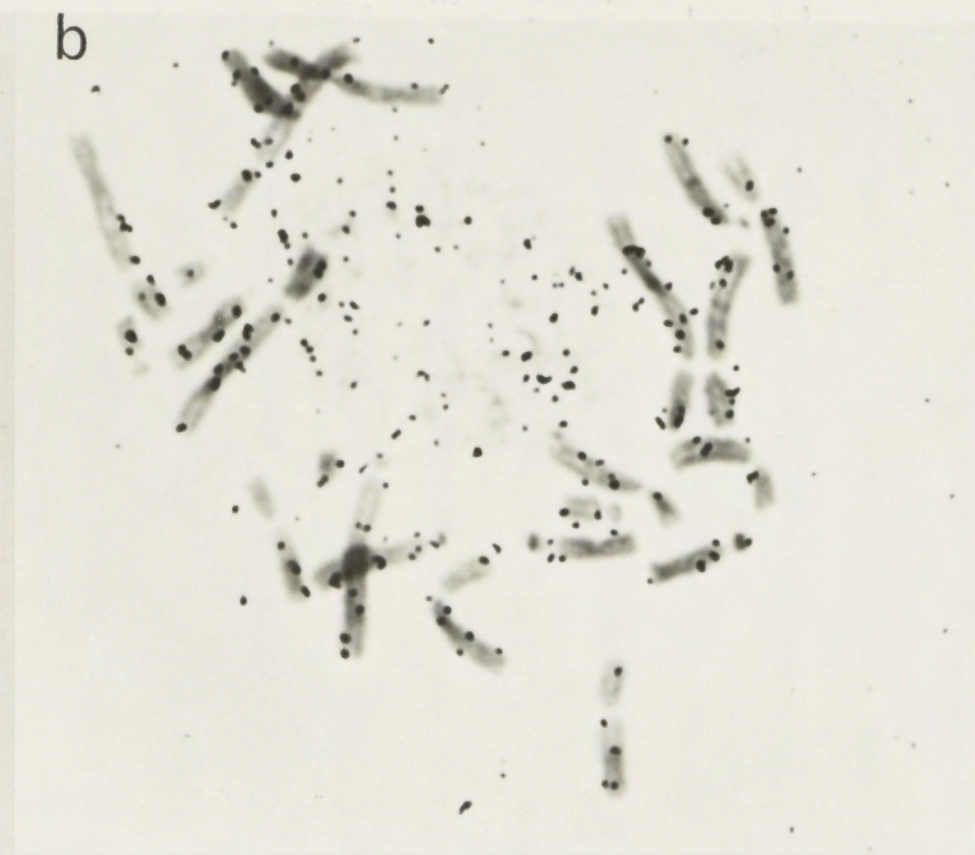
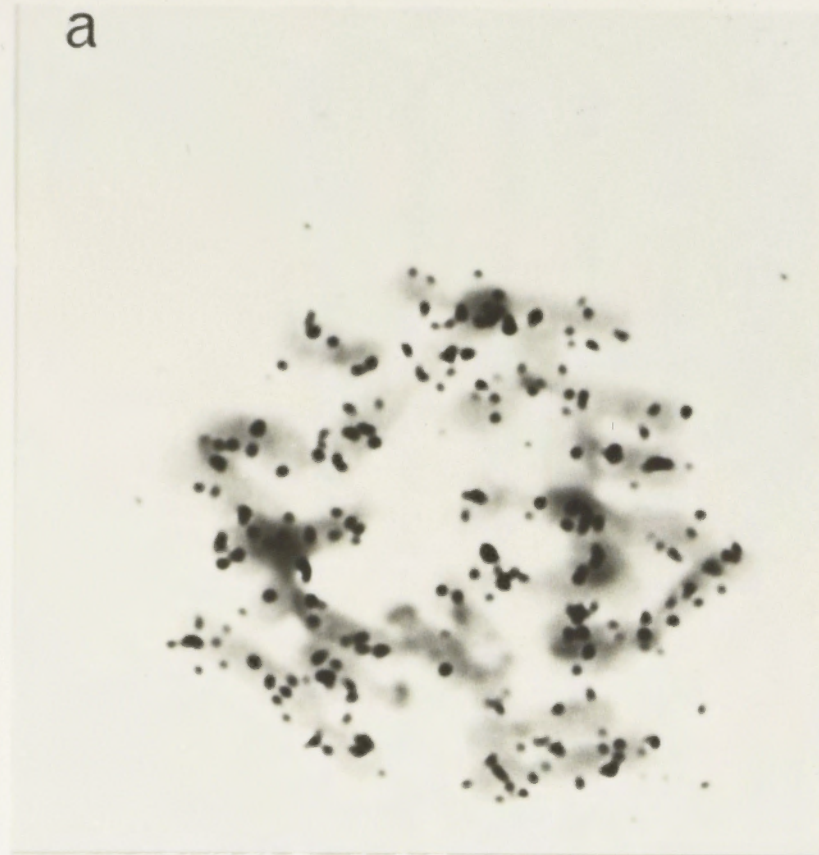


Figure 6-2. In situ hybridization utilizing the Eco RI fragments. a.  $^3\text{H}$  c-RNA from pGEA hybridized to a Peregrine Falcon (Moreton) individual (7 day exposure); b.  $^3\text{H}$  c-RNA from pGEB hybridized to a Peregrine Falcon individual (7 day exposure); c.  $^3\text{H}$  c-RNA from pGEC hybridized to a Peregrine Falcon individual (7 day exposure); d.  $^3\text{H}$  c-RNA from pGEC hybridized to a C. species nova 1 individual (4 day exposure).







various Caledia taxa. However, whether or not these three sequences are always closely associated in the genome, as would be suggested by their presence in the single clone, is not known.

Partial sequence data for the three repeats is given in Figure 6-3. Computer comparison of these sequences has not revealed any extended regions of homology greater than 9 bp. However, two observations are worth comment. First, since only a portion of the repeats have been sequenced, there may be extensive regions of homology which have not been identified. Second, although there are no regions of significant homology between the three repeats, there are specific motifs which are shared between them. For example, the nucleotide sequences possess poly-CA subrepeats. This is most obvious in the pGEA repeat which has a (CA)<sub>25</sub> sequence (Fig. 6-4). Such an extensive region of alternating purines and pyrimidines might form left-handed or Z-DNA (Nordheim et al., 1981). The potential structural and functional attributes of such DNA, in light of the Caledia Eco sequences, are discussed in the final chapter.



**Figure 6-3.** Partial nucleotide sequences for the three dispersed repeats. 1. pGEC, 2. pGEA and 3. pGEB.

**Note:** One half of the Eco RI recognition site (TTC) has been included as the starting point for each of the three repeats.

1            10            20            30            40            50            60            70            80            90  
:            :            :            :            :            :            :            :            :

1 TTCATTGRCCCAGGAATGGAAAACCTTTATATACACATTCTGGGGTCAGATACAACACATGATCACACAGACAGAACCATAGGCACATAAG  
2 TTCTCAAGGAATGGATTGTTGTTTATTACGAAGACGAGGCAACCATTATAAAAACAACAAAAATGTTTCAGACGTTTGTAGTCCCAAGGGT  
3 TTCCAAAACCACAACCCAGTGATACAAAATGCTCATCTTATGAAAATGTACTGACCCAGACATGAAACATGGACTATGTAACAATGTAGT

100            110            120            130            140            150            160            170            180  
:            :            :            :            :            :            :            :

1 ACACAGGCAACAGAGCATGCACATGTCGGCACTAGTACCGTGTATATCTCTTTTCGCAGCAATGCAGGCTGCATTCTCCATGGAGAGGATC  
2 CCCTGATTAGTTCATCGGTTTCTAGACTTACAACCTACCTGAACTAAATTACGCTGGACAATACACACACACACACACACACACACAC  
3 TGATCCTAAGGCTGATTTGATGCAACTATCCACCGTACCTTGTCCTGTGCAAGTTCTTCATCTCTGAATATGATGACACCTATATCCATT

190            200            210            220  
:            :            :            :

1 GTAGAGGTGCTGGATGTAGTCCTGTGGACGGCATGCAAGCATTT  
2 ACACACACACACACACACACAGCCATGCCG  
3 CTTATT



Figure 6-4. Results from sequencing experiment with the clone pGEA showing the (CA)<sub>25</sub> region. This region occurs from base position 154 to 203 (Fig. 6-3).

A C G T





## CHAPTER 7: FINAL REMARKS

### 7.1 Introduction

In Chapter 1 (Section 1.1) I proposed five questions to be addressed in this thesis. The first four related to how highly repeated DNA sequences change through time and the last dealt with the aspect of function. In this regard, each of these questions have been answered to varying degrees in Chapters 3-6 and Appendix I. In this final chapter I will consider the two major themes of these questions; the evolution and putative functionality of highly repeated sequences.

### 7.2 Evolution of Highly Repeated DNA Families

The study of highly repeated sequence evolution, like comparable chromosomal, allozymic and morphological studies, has followed two main paths. The approaches which have been used include 1) the comparison of sequence families between closely and more distantly related taxa (e.g., see Sutton and McCallum, 1972; Barnes et al., 1978) and 2) the analysis of intraspecific variation between different satellites (e.g., see Gall and Atherton, 1974; Peacock et al., 1977). These different approaches led to the observation that a specific highly repeated sequence family had less intraspecific than interspecific variation. This phenomenon was labelled "concerted evolution" (Zimmer et al., 1980) because the multigene and satellite repeat units seemed to evolve in concert. This aspect of highly repeated DNA has been noted in studies involving a number of sequence families, and led to the conclusion that there was intraspecific sequence homogeneity within repeat families (e.g.,

Appels et al., 1981; Coen et al., 1982; Strachan et al., 1982). Thus, Strachan et al. (1982) found that a majority of repeats from the Drosophila "500" and "360" highly repeated sequence families were characterized by certain restriction endonuclease cleavage sites and patterns of genomic organization that were diagnostic for all species with the exception of D. mauritiana and D. simulans, the two most closely related species.

Similarly, the "480 bp" family repeat isolated from the rye genome was described as relatively homogeneous on the basis of endonuclease digestion and T<sub>m</sub> analysis (Appels et al., 1981), and it was estimated that there was a maximum of 2% divergence between the members of this repeat family. In contrast, the data from the Caledia study presented in this thesis, as well as the recent findings concerning other repeat families (Trick and Dover, 1984; Strachan et al., 1985; Appels and Moran, in prep.), suggest that the concept of "homogeneity" needs to be qualified in terms of the large amount of variation that is consistently present within a sequence family. Therefore, although there is evidence that at least a portion of the C. captiva 168 bp sequence is relatively free of mutations (see below), it is apparent that much of the sequence is free to incorporate changes. For example, direct nucleotide sequencing of repeats from this family has shown that there can be as much as 10% sequence variation between repeats from a single individual. Furthermore, the Bam family sequences, presented in Chapter 5, demonstrate even greater levels of variation when equivalent intertaxonomic comparisons are made.



If maintenance of any portion of a repeated sequence is occurring, this has implications concerning how sequences evolve within the genome. Arnold et al. (1985a; Chapter 3) have demonstrated that an unchanged 18 bp region within the 168 bp repeat is not predicted on the basis of a statistical test developed by Brown and Clegg (1983). Similarly, it has been shown that a 27 bp sequence that is invariant in clones of 180 bp repeats from maize, teosinte and Tripsacum is statistically significant (Brown and Clegg, 1983; Dennis and Peacock, 1984). What then are the processes that account for this pattern of evolution? It has been suggested that the evolution of a highly repeated sequence family would necessarily have a number of stages (Dover, 1982; Arnheim, 1983). Firstly, the ancestral sequence would presumably arise by chance at a single chromosomal site, and then be amplified to some degree. The alternative to this would be that the sequence originates in a number of chromosomal locations and is then amplified. This is less likely given parsimony, however, it remains a possibility. If the sequence arises at a single site, it may or may not be disseminated to different locations in the genome. The 168 bp family from C. captiva shows a wide range of variation in its chromosomal distribution (Arnold and Shaw, 1985; Chapter 4). The S.E.A. and Moreton taxa have numerous chromosomal locations that show homology to a probe synthesized from this family. These sites include distal, interstitial and proximal positions. The Southern, Northern and Papuan Torresian races, on the other hand, are located at telomeric sites on chromosomes 10, 11 and 12, chromosome 11 and chromosomes 6 and 12, respectively. Thus, the

168 bp sequence, having evidently originated in the progenitor of the C. captiva taxa, has been distributed to from one (N. Torresian) to all 24 (Moreton and S.E.A.) of the chromosomes in the complement. Included in this spread was a differential amplification in the various taxa; the Torresian taxon has approximately 40,000 copies and the Moreton/S.E.A. taxa possess 150,000 copies of this repeat. In contrast, the sequences within each genome may maintain a degree of "communication" in that each repeat demonstrates a level of similarity which does not appear to be explained on the basis of random chance (ie., the 18 bp conserved region). Likewise, Coen et al. (1982) have noted the differential abundance of certain non-transcribed spacer length variants in the rDNA family of D. melanogaster. From this, they concluded that the spacer length variants could be differentially amplified throughout an array of rDNA genes.

Specific mechanisms have been suggested to account for the spread of sequence variants, not only in multigene families, but also in highly repeated sequence families. In a theoretical analysis of unequal crossingover, it was demonstrated that the majority of newly arisen sequence variants would be lost through this process (Smith, 1976). However, by chance a variant could infrequently be amplified to the extent that it would become the major sequence type. Experimental analyses of yeast rDNA genes have shown that unequal crossingover between sister chromatids can occur in mitosis and meiosis (Szostak and Wu, 1980; Petes, 1980). However, although it is assumed that this type of mechanism is at least partly responsible for both the



amplification and the reduction of highly repeated sequence families, no definitive demonstration of this has been put forward to date. The second mechanism that has been suggested to account specifically for the distribution of repeats between different chromosomes is transposition. The occurrence of moderately repetitive sequences that have the ability to transpose themselves from one location to another has been well documented (see section 7.3 for discussion). The potential for the movement of other repeated sequences along with the transposon is obvious. However, as with unequal crossingover, the involvement of these sequences in the spread of highly repeated sequence families has yet to be demonstrated. Finally, a form of gene conversion has been implicated in the evolution of sequence families. As discussed in Chapter 1, a gene conversion event involves the changing of one allele at a locus into a different allelic state (Roman, 1963). This process may have been responsible for the intergenic conversion events between genes in the chorion and globin multigene families (Slightom et al., 1980; Iatrou et al., 1984). Thus, although the Hc-A and Hc-B chorion genes examined by Iatrou et al. (1984) originate from different gene families, they demonstrate long stretches of sequence homology in their 3' untranslated regions. This homology could be due to the transfer of sequence characteristics from one gene to the other. Similarly, a conversion event has been suggested as an explanation for the similarity between two "nonallelic" genes coding for human fetal globin (Slightom et al., 1980). The region of homology occurs within an intervening sequence and is striking because the "allelic" genes differ

markedly in this same region.

It is apparent from the above, as well as the discussion in Chapter 1, that certain processes have been implicated in the evolution of highly repeated sequences. However, it is just as apparent that, as of yet, the available data merely infer the action of these processes in the evolution of such sequence families. This inherent weakness is present in all evolutionary studies in that the separation and analysis of the causal factors is extremely difficult. It is feasible that, in future analyses, highly repeated sequences could be introduced into a genome via a transposable element. Thus inserted, the sequence could be monitored at a molecular level to test for subsequent amplification and/or transposition. In addition it might be possible, via hybridization, to introduce a sequence such as the 144 bp repeat from Papuan Torresian into the Northern or Southern Torresian genomes. Since this sequence is present in high copy number in the Papuan Torresian genome it might become similarly amplified in these novel genomic backgrounds. With such an approach one could predict the frequency and pattern of incorporation of repeat sequences into the genome.

### 7.3 Functional Aspects of Highly Repeated Sequences

It is apparent that any discussion of putative functional aspects of highly repeated DNA must recognize that an as yet unquantified proportion of these DNAs may have no function (Orgel and Crick, 1980; Doolittle and Sapienza, 1980; Miklos, 1985). In point of fact, the current consensus pertaining to highly repeated sequences may have been expressed by Smith (1976). His



theoretical examination of the evolution of highly repeated sequences via unequal crossover events led him to the conclusion that "... a pattern of tandem repeats is the natural state of DNA whose sequence is not maintained by selection." However, the difficulty in ascribing or denying any function to a repeated DNA family has been highlighted by studies of DNA in the classes of intervening sequences and transposable elements. Darnell (1978) has pointed out that nontranscribed intervening sequences, at one stage of viral development, are at another time transcribed. Therefore, the production of mRNA by the virus AD2 at a relatively later time (post-infection) includes all of the earlier transcribed sequences in addition to mRNAs which account for almost all of the remainder of the viral genome (Flint, 1977). In contrast, it is probable that some eukaryotic intervening sequences may never be transcribed nor have a regulatory function (Darnell, 1978). This has led to the suggestion that a "genes-in-pieces" structure is the primitive genomic trait, in contrast to the more highly "streamlined" form found in bacteria and viruses (Doolittle, 1978). The intervening sequences may allow a certain degree of evolutionary plasticity in terms of evolving new proteins via novel splicing of the mRNA, if the new protein produces a selective advantage (Gilbert, 1978). However, it is also possible that "... many eukaryotic cells simply do not have an equivalent facility for deleting "excess" sequences so that nonfunctional "intervening" sequences would be maintained for a longer time in eukaryotes" (Darnell, 1978).

Transposable elements in the genus Drosophila were initially described due to their ability to cause spontaneous mutations (Green, 1977). Subsequently, these sequences were demonstrated to be repetitious and to have the ability to excise and insert in numerous genomic locations (Ilyin et al., 1978; Potter et al., 1979; Strobel et al., 1979). The phenotypic effects produced by these elements may be directly attributable to their insertion into either regulatory regions or within structural genes (Rubin, 1983). These elements may increase genetic variability through their effecting mutations and chromosomal rearrangements as well as preventing the admixture of certain gene pools through hybrid dysgenesis (Kidwell et al., 1977). However, whether or not these factors are of selective importance to the organism is not yet known (Rubin, 1983). These several examples illustrate the difficulty in assigning or dismissing any functional attribute for a specific sequence family, however, it is interesting that the majority of the putative functions for highly repeated sequences (ie., nucleosome phasing, Strauss and Varshavsky, 1984; microtubule-chromosome interaction, Avila et al., 1983; influence on gene activity, e.g., Moss, 1983; Andrews et al., 1984; chromosomal attachment to the nuclear envelope, Rae and Franke, 1972; and redistribution of chiasmata, Miklos and Nankivell, 1976) involve protein-DNA interactions.

In Drosophila, spontaneous or induced chromosomal rearrangements that result in the alignment of normally separated heterochromatic and euchromatic segments, may produce profound changes in phenotype. The phenotypic expression caused by such cytological reorganizations were first defined by Muller (1930)



and were subsequently named "variegated position effects" due to their mosaic expression in the mutant individual (Lewis, 1950). This effect is apparent both when euchromatic loci are positioned within heterochromatin (e.g., Ratty, 1954) and when loci normally located in heterochromatic regions are transferred to a distal location relative to the heterochromatin (e.g., Parker, 1967). A number of factors impinge upon the degree to which the phenotypic abnormalities associated with this phenomenon are expressed. For example, although variegation is extensively induced by autosomal heterochromatin, it has been suggested that greater than 50% of the rearrangements of X and Y heterochromatin to euchromatic regions have no effect on the euchromatic loci (Spofford, 1976). Furthermore, by manipulating additional environmental and genetic factors, one can lessen the expression of the mutant phenotype. Such an effect is produced by altering the temperature at which the flies are reared. Therefore, Gowan and Gay (1934) described a greater level of mutant regions in the eyes of individuals with the "Mottled-3" phenotype when they were reared at lower temperatures relative to those flies reared at higher temperatures.

A further consequence of the variegated position effect seems to suggest at least two possible regulatory mechanisms. Cohen (1962) examined the effects of a rearrangement that placed heterochromatin and the centromere on opposite sides of several marker genes. She found that both bodies of heterochromatin were able to affect not only the loci proximal to them, but also each of the other more distally located genes. This so-called

"spreading effect" indicates that the mechanism responsible for the phenotypic change originates in the heterochromatin and moves into the genic region (Cohen, 1962). It has been suggested that the heterochromatic region has a specific property that determines its highly condensed, and presumably transcriptionally inactive, nature. Furthermore, it is assumed that the normal junction sites between this chromatin and the euchromatin have the ability to limit the extent of this condensation (Spofford, 1976). However, when a new junction is formed, the heterochromatin most likely causes a defect in the transcriptional process that may well be due to the production of structural changes in the euchromatin (Rushlow et al., 1984). Thus, the condensation of the euchromatin into a heterochromatin-like structure might render the gene loci transcriptionally inactive. This type of structural change has seemingly been cytologically detected in that the variegating loci lack their normal banding patterns in Drosophila polytene chromosomes and appear more like the proximal heterochromatin (Hartmann-Goldstein, 1967). A molecular analysis of three white mutants, that demonstrate variegated position effects, led Tartof et al. (1984) to propose a model in which there are initiation and termination sites that establish the boundaries of the heterochromatized DNA. Furthermore, if these sites exist, they are most likely infrequently distributed within the heterochromatin (Tartof et al., 1984). Therefore, the spreading effect and the entire phenomenon of variegating position effect may be explained by such factors via their effect on the heterochromatization of a portion of the proximal euchromatin.



A second explanation for the observed position effect relates to the findings of Moss (1983) and Andrews et al. (1984). These authors have shown that Xenopus repetitive sequence elements, related to promoter sequences, can enhance or suppress the activity of the respective promoter. In the situation described by Andrews et al. (1984) this occurs when a satellite sequence is placed into a plasmid containing the 5S genes and when the repeated sequence is present in a separate plasmid vector. This effect was produced when the plasmids carrying the sequences were either injected into oocyte nuclei or incubated in HeLa S-100 extracts. The competition was attributed to the presence of a RNA polymerase III "transcription unit" in the satellite repeat that was 10-25-fold more efficient than that of the 5S RNA genes. Therefore, these satellite sequences have the capability of profoundly effecting the transcription of certain essential genes. If these sequences were "competitive" throughout development, the organism might be detrimentally affected. Furthermore, it appears that, in vivo, most of the Xenopus satellite I sequence family is in non-transcribed regions of the chromatin (Jamrich et al., 1983). This finding has led to the suggestion that these sequences may not bind a significant number of transcription factors (Andrews et al., 1984). An alternative to this explanation is that such sequences do have the ability to bind transcription factors and, therefore, may affect the process of gene regulation (Appels and Peacock, 1978). However, if termination sequences are present proximal to the RNA polymerase binding site, one would not predict the production of

transcripts of these "competitor" sequences. If the movement of, for instance, the Xenopus satellite sequence to a euchromatic region caused the previously transcriptionally inert sequence to be available to bind polymerase molecules, the genes proximal to such a highly repeated sequence could be inactivated as a result of a competitive interaction. Therefore, the pretranslational effects detected by Rushlow et al. (1984) for the rosy locus might be the result of the positioning of a competitively superior sequence (in terms of binding necessary transcriptional molecules) near this locus.

Findings from a study of a developmental transition in Xenopus embryos indirectly suggest the potential importance of quantitative variation in satellite DNA (Newport and Kirschner, 1982a, b). It was discovered that at the midblastula stage, there is a major change in development which includes the initiation of transcription. The factor which apparently governs this change is the ratio of the nucleus to the cytoplasm (Newport and Kirschner, 1982a). Furthermore, it was discovered that the introduction of plasmid DNA equivalent to the amount present in the cell stage just prior to the developmental transition induced premature transcription (Newport and Kirschner, 1982b). In other words, the developmental switch could be prematurely initiated through the addition of excess DNA. These authors suggested a model, whereby, the initiation of transcription was governed by the titration of "suppressor molecules" which bound to the DNA. When all of the molecules were associated with DNA, the remaining unbound DNA was free to act as a template for transcription.

In relation to the above studies, the data presented for the



Caledia highly repeated sequences demonstrate that at least some of the repeat families possess structural features that have a potential for involvement in the organization and, thereby, the functioning of the genome. For example, an interesting characteristic of the 168 bp repeats, involving the 18 bp conserved region is the presence of two-fold or dyad symmetry. This type of symmetry was first found by Gilbert and Maxam (1973) in the nucleotide sequence of the lac operator. These authors suggested that the dyad symmetry present in this region might facilitate the interaction of the repressor protein (known to bind this sequence) with the DNA on a two-fold symmetrical axis. The two 6 bp stretches at either end of the lac operator region of symmetry are separated by a distance of approximately one turn of the helix (Gilbert and Maxam, 1973) and, therefore, are located on the same side of the DNA molecule.

As shown in Chapter 3 the major area of dyad symmetry in the 168 bp repeat involves 17 bp of the 18 bp conserved region and 25 bp immediately adjacent to this stretch. The similarity of this region to not only the protein binding site of the lac operator, but also to other sites of protein-DNA interaction (Dickson et al., 1975; Bennett et al., 1978; Bennett and Yanofsky, 1978), is quite striking. The major difference between the area of dyad symmetry in the 168 bp repeat and the previously described protein binding sites is that of length. The area of symmetry, which has as one "arm" the conserved region, is 42 bp in length. This is exactly twice the length of the lac operator site (Gilbert and Maxam, 1973), over twice the size of the trp

operator binding site (Bennett et al., 1978; Bennett and Yanofsky, 1978) and corresponds to exactly four turns of the DNA helix. It is also of interest to note that when the entire 168 bp repeat is examined for the presence of two-fold symmetry that a region beginning at position 22 and ending at position 147 shows extensive dyad symmetry. This stretch corresponds to exactly 12 turns of the helix and once again this is suggestive of a region available for protein-DNA interaction.

The two-fold symmetry present in the Caledia sequence is not unique for highly repeated DNA. Hsieh and Brutlag (1979a) described a protein which preferentially bound to a 359 bp Drosophila satellite sequence. By utilizing restriction enzyme analysis of the DNA-protein complex it was demonstrated that only a small portion of the repeat was involved in binding the protein (Hsieh and Brutlag, 1979a). An examination of the nucleotide sequence of the 359 bp repeat revealed an area, corresponding to the general region of the DNA-protein binding site, which possessed dyad symmetry (Hsieh and Brutlag, 1979b).

If the regions of symmetry in the Caledia repeat, in particular the 42 bp stretch, function as protein binding sites it is necessary to accomodate any variability in base pair sequence among the repeats. Although the 42 bp region contains 17 bases which are highly conserved there are also 14 variable sites. These represent single base pair changes involving either substitutions or additions. It is important to note that the frequency of change at these sites does not exceed 0.06 (see Chapter 3, Figs. 5 and 6). In addition, Bennett and Yanofsky (1978) have demonstrated that although point mutations within the



trp operator produce constitutive mutants, these mutants have levels of constitutivity ranging from 20% to 70%. This would seem to indicate that while the repressor-operator interaction is indeed hindered by point mutations, it is not completely excluded. Therefore, one might predict that the point mutations present in the putative recognition site in the Caledia repeats would not necessarily negate the potential for formation of a specific protein-DNA complex. It is also possible that the highly conserved nature of the 18 bp region reflects a necessary property of having at least one half of the protein-DNA interaction site invariant.

A second novel structure detected in the Caledia highly repeated sequences was the  $d(CA)_{25}$  tract in one of the Eco repeats. The alternating pyrimidine/purine format of this sequence is significant in that this is believed to be the basic structure necessary to form Z-DNA (Nordheim et al., 1981; Felsenfeld et al., 1982). Z- or left-handed DNA was initially detected by X-ray crystallographic analysis of a  $d(CG)_3$  fragment (Wang et al., 1979). This DNA sequence was demonstrated to form a double helical crystal like B-DNA, however, a number of major structural differences were apparent between the two DNA forms. For example, the Z-DNA "helix sense" was left- rather than right-handed, there were 12 as opposed to 10 residues per helical turn and the diameter was approximately 18 angstroms in contrast to the 20 angstrom diameter of the B-DNA helix (Wang et al., 1979). The presence of the Z-DNA structure has been detected in the interband regions of Drosophila melanogaster polytene chromosomes

(Nordheim et al., 1981). Although the occurrence of this conformation in vivo remains equivocal (Hill and Stollar, 1983; Robert-Nicoud et al., 1984), Nordheim et al. (1981) have noted the amenability of the Z-DNA conformation to a gene regulatory role due to the reversibility of the structure. Furthermore, they suggested that not only may the Z-DNA alter the proximal DNA structure, but it may, through supercoiling, affect the transcription of gene regions distal to itself. Such a speculation indicates the potential significance for the unique structure of the Eco repeats and their dispersed nature in the Caledia genome.

The findings from all of the studies discussed above, including those concerning Caledia, suggest an exciting dimension to highly repeated DNA. It is evident that the general lack of transcriptional activity, late-replication, cytological appearance and the unique molecular structure of tandemly repeated sequences are not necessarily, as assumed by many authors, indicative of genetic inertness. Rather, they may well belie profound and complex functional attributes within an organism (John and Miklos, 1979). In relation to the putative role of competition, discussed above, there are two essential components which must be present. First, the "pool" of protein molecules (eg., suppressors, initiators, cofactors) must be limited in their amount, leading to an effect through titration. Second, the DNA-protein binding reaction must be reversible, resulting in a state of equilibrium. Thus, a perturbation in the numbers of unbound protein factors available for binding would result in an overall effect at the level of the chromosome. Many aspects of



cell metabolism could effect these two components. One possible parameter is the addition or deletion of competitor repetitive DNA sequences. In this regard, the Caledia system offers the very real potential to resolve how specific highly repeated DNA families interact with other components within the cell. 1) Studies of the potential for amplification and spread of repeated sequences can be made utilizing the techniques of tissue culture and cell fusion. Thus, sequences unique to one cytotype could be introduced into another complement and any subsequent evolution of this sequence could be monitored. 2) Furthermore, as demonstrated by Andrews et al. (1984), satellite sequences can be tested for any effect they may have on gene transcription through the use of competition experiments. Such determinations should also be informative for the Caledia highly repeated sequence families. 3) Finally, it is currently possible to test for the binding of molecules to specific DNA regions in the chromatin matrix. Church and Gilbert (1984) have described a technique that allows an assay of the structure of uncloned DNA in terms of its interaction with other molecules. Therefore, this method should allow an initial assessment of whether the area of dyad symmetry, in the 168 bp repeats, is involved in the binding of proteins. It is apparent that whatever the results of such experiments, these data will enhance our understanding of the in vivo properties and evolution of highly repeated sequences.

BIBLIOGRAPHY. Cumulative for Chapters 1 through 7 and Appendix I.

- Andrews, D.L., L. Millstein, B.A. Hamkalo and J.M. Gottesfeld.  
1984. Competition between Xenopus satellite I sequences and Pol III genes for stable transcription complex formation.  
N.A.R. 12: 7753-7769.
- Appels, R. 1982. The molecular cytology of wheat-rye hybrids.  
Int. Rev. Cytol. 80: 93-132.
- Appels, R., E.S. Dennis, D.R. Smyth and W.J. Peacock. 1981. Two repeated DNA sequences from the heterochromatic regions of rye (Secale cereale) chromosomes. Chromosoma 84: 265-277.
- Appels, R., C. Driscoll and W.J. Peacock. 1978. Heterochromatin and highly repeated DNA sequences in rye (Secale cereale). Chromosoma 70: 67-89.
- Appels, R. and J. Dvorak. 1982. The wheat ribosomal DNA spacer region: its structure and variation in populations and among species. Theor. Appl. Genet. 63: 337-348.
- Appels, R. and L. Moran. 1984. Molecular analysis of alien chromatin introduced into wheat. Pp. 529-557 in J.P. Gustafson, ed. 16th Stadler Genetics Symposium, Gene Manipulation in Plant Improvement. Plenum Press, New York.
- Appels, R. and W.J. Peacock. 1978. The arrangement and evolution of highly repeated (satellite) sequences with special reference to Drosophila. Int. Rev. Cytol. Suppl. 8: 69-126.



- Arnheim, N. 1983. Concerted evolution of multigene families. Pp. 38-61 in M. Nei and R.K. Koehn, eds. Evolution of genes and proteins. Sinauer Associates, Inc., Sunderland, Mass.
- Arnold, M.L., R. Appels and D.D. Shaw. 1985a. The heterochromatin of grasshoppers from the Caledia captiva species complex. I. Sequence evolution and conservation in a highly repeated DNA family. Mol. Biol. Evol., in press.
- Arnold, M.L., R. Appels and D.D. Shaw. 1985b. Evolution and conservation in a highly repeated DNA family from the grasshopper Caledia captiva. Cytobios, in press.
- Arnold, M.L. and D.D. Shaw. 1985. The heterochromatin of grasshoppers from the Caledia captiva species complex II. Cytological organization of tandemly repeated DNA sequences. Chromosoma, in review.
- Avila, J., E. Montejó de Garcini, F. Wandosell, A. Villasante, J.M. Sogo and N. Villanueva. 1983. Microtubule - associated protein MAP<sub>2</sub> preferentially binds to a dA/dT sequence present in mouse satellite DNA. EMBO J. 2: 1229-1234.
- Bankier, A.T. and B.G. Barrell. 1983. Shotgun DNA sequencing. Pp. 1-34 in R.A. Flavell, ed. Techniques in nucleic acid biochemistry. Elsevier, Amsterdam.
- Barnes, S.R., D.A. Webb and G. Dover. 1978. The distribution of satellite and main-band DNA components in the melanogaster species subgroup of Drosophila I. Fractionation of DNA in actinomycin D and distamycin A density gradients. Chromosoma 67: 341-363.

- Beattie, W.G. and D.M. Skinner. 1972. The diversity of satellite DNAs of Crustacea. *Biochim. Biophys. Acta* 281: 169-178.
- Bennett, G.N., M.E. Schweingruber, K.D. Brown, C. Squires and C. Yanofsky. 1978. Nucleotide sequence of the promoter-operator region of the tryptophan operon of Escherichia coli. *J. Mol. Biol.* 121: 113-137.
- Bennett, G.N. and C. Yanofsky. 1978. Sequence analysis of operator constitutive mutants of the tryptophan operon of Escherichia coli. *J. Mol. Biol.* 121: 179-192.
- Biro, P.A., A. Carr-Brown, E.M. Southern and P.M.B. Walker. 1975. Partial sequence analysis of mouse satellite DNA: evidence for short range periodicities. *J. Mol. Biol.* 94: 71-86.
- Bishop, J.O., F.W. Robertson, J.A. Burns and M. Melli. 1969. Methods for the analysis of deoxyribonucleic acid - ribonucleic acid hybridization data. *Biochem. J.* 115: 361-370.
- Bolton, E.T., R.J. Britten, D.B. Cowie, R.B. Roberts, P. Szafranski and M.J. Waring. 1965. Interactions of Nucleic Acids. *Carnegie Inst. Wash. Year Book* 64: 313-341.
- Borgmeyer, U., J. Nowock and A.E. Sipple. 1984. The TGGCA-binding protein: a eukaryotic nuclear protein recognizing a symmetrical sequence on double-stranded linear DNA. *N.A.R.* 12: 4295-4311.



- Bostock, C.J., S. Christic and I.J. Lauder. 1976. S-phase patterns of replication of different satellite DNAs in three species of Dipodomys (Kangaroo Rat). J. Mol. Biol. 108: 417-433.
- Botchan, M.R. 1974. Bovine satellite I DNA consists of repetitive units 1400 base pairs in length. Nature 251: 288-292.
- Britten, R.J. and D.E. <sup>Kohne</sup>. 1967. Repeated nucleotide sequences. Carnegie Inst. Wash. Year Book 66: 73-88.
- Britten, R.J. and D. E. Koehn. 1968. Repeated sequences in DNA. Science 161: 529-540.
- Brown, A.H.D. and M.T. Clegg. 1983. Analysis of variation in related DNA sequences. Pp. 107-132 in B.S. Weir, ed. Statistical analysis of DNA sequence data. Marcel Dekker, Inc., New York.
- Brown, D.D., P.C. Wensink and E. Jordan. 1971. Purification and some characteristics of 5S DNA from Xenopus laevis. P.N.A.S. 68: 3175-3179.
- Brutlag, D.L. 1980. Molecular arrangement and evolution of heterochromatic DNA. Ann. Rev. Genet. 14: 121-144.
- Brutlag, D., R. Appels, E.S. Dennis and W.J. Peacock. 1977. Highly repeated DNA in Drosophila melanogaster. J. Mol. Biol. 112: 31-47.
- Brutlag, D.L. and W.J. Peacock. 1975. Sequences of highly repeated DNA in Drosophila melanogaster. Pp. 35-45 in W.J. Peacock and R.D. Brock, eds. The Eukaryote Chromosome. Australian National University Press, Canberra.

- Case, M.E. and N.H. Giles. 1964. Allelic recombination in Neurospora: tetrad analysis of a three-point cross within the pan-2 locus. Genetics 49: 529-540.
- Church, G.M. and W. Gilbert. 1984. Genomic sequencing. P.N.A.S. 81: 1991-1995.
- Coates, D.J. and D.D. Shaw. 1982. The chromosomal component of reproductive isolation in the grasshopper Caledia captiva. I. Meiotic analysis of chiasma distribution patterns in two chromosomal taxa and their F<sub>1</sub> hybrids. Chromosoma 86: 509-531.
- Coates, D.J. and D.D. Shaw. 1984. The chromosomal component of reproductive isolation in the grasshopper Caledia captiva. III. Chiasma distribution patterns in a new chromosomal taxon. Heredity 53: 85-100.
- Coen, E., T. Strachan and G. Dover. 1982. Dynamics of concerted evolution of ribosomal DNA and histone gene families in the melanogaster species subgroup of Drosophila. J. Mol. Biol. 158: 17-35.
- Cohen, J. 1962. Position-effect variegation at several closely linked loci in Drosophila melanogaster. Genetics 47: 647-659.
- Cohen, E.H. and S.C. Bowman. 1979. Detection and location of three simple sequence DNAs in polytene chromosomes from virilis group species of Drosophila. Chromosoma 73: 327-355.
- Corneo, G. 1976. Do satellite DNA's function as sterility barriers in eukaryotes? Evol. Theory 1: 261-265.



- Daly, J.C., P. Wilkinson and D.D. Shaw. 1981. Reproduction isolation in relation to allozymic and chromosomal differentiation in the grasshopper Caledia captiva. Evolution 35: 1164-1179.
- Darnell, J.E., Jr. 1978. Implications of RNA-RNA splicing in evolution of eukaryotic cells. Science 202: 1257-1260.
- Denhardt, D.T. 1966. A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Comm. 23: 641-646.
- Dennis, E.S. and W.J. Peacock. 1984. Knob heterochromatin homology in maize and its relatives. J. Mol. Evol. 20: 341-350.
- Dickson, R.C., J. Abelson, W.M. Barnes and W.S. Reznikoff. 1975. Genetic regulation: The lac control region. Science 187: 27-35.
- Donehower, L., C. Furlong, D. Gillespie and D. Kurnit. 1980. DNA sequence of baboon highly repeated DNA: evidence for evolution by nonrandom unequal crossovers. P.N.A.S. 77:2129-2133.
- Donehower, L. and D. Gillespie. 1979. Restriction site periodicities in highly repetitive DNA of primates. J. Mol. Biol. 134: 805-834.
- Doolittle, W.F. 1978. Genes in pieces: where they ever together? Nature 272: 581-582.
- Doolittle, W.F. and C. Sapienza. 1980. Selfish genes, the phenotype paradigm and genome evolution. Nature 284: 601-603.

- Dover, G. 1982. Molecular drive: a cohesive mode of species evolution. *Nature* 299: 111-116.
- Dromey, R.G. 1979. A fast algorithm for text comparison. *Aust. Comp. J.* 11: 63-67.
- Dunsmuir, P. 1976. Satellite DNA in the kangaroo Macropus rufogriseus. *Chromosoma* 56: 111-125.
- Dvorak, J. and R. Appels. 1982. Chromosome and nucleotide sequence differentiation in genomes of polyploid Triticum species. *Theor. Appl. Genet.* 63: 349-360.
- Endow, S.A., M.L. Polan and J.G. Gall. 1975. Satellite DNA sequences of Drosophila melanogaster. *J. Mol. Biol.* 96: 665-692.
- Epstein, D.A., F.R. Witney and A.V. Furano. 1984. The spread of sequence variants in Rattus satellite DNAs. *N.A.R.* 12: 973-988.
- Felsenfeld, G., J. McGhee, D.C. Rau, W. Wood, J. Nickol and M. Behe. 1982. Chromatin conformation and gene activity. Pp. 121-135 in B.W. O'Malley, ed. *Gene Regulation*. Academic Press, New York.
- Flamm, W.G., H.E. Bond, H.E. Burr and S. B. Bond. 1966. Satellite DNA isolated from mouse liver; some physical and metabolic properties. *Biochim. Biophys. Acta.* 123: 652-654.
- Flamm, W.G., M. McCallum and P.M.B. Walker. 1967. The isolation of complementary strands from a mouse DNA fraction. *P.N.A.S.* 57: 1729-1734.



- Flavell, R.B., M.D. Bennett, J.B. Smith and D.B. Smith. 1974. Genome size and the proportion of repeated nucleotide sequence DNA in plants. *Biochem. Genet.* 12: 257-269.
- Flint, J. 1977. The topography and transcription of the adenovirus genome. *Cell* 10: 153-166.
- Franke, W.W. 1974. Structure, biochemistry, and functions of the nuclear envelope. *Int. Rev. Cytol. Suppl.* 4: 71-236.
- Fry, K., R. Poon, P. Whitcome, J. Idriss, W. Salser, J. Mazrimas and F. Hatch. 1973. Nucleotide sequence of HS- $\beta$  satellite DNA from kangaroo rat Dipodomys ordii. *P.N.A.S.* 70: 2642-2646.
- Fry, K. and W. Salser. 1977. Nucleotide sequences of HS- $\alpha$  satellite DNA from kangaroo rat Dipodomys ordii and characterization of similar sequences in other rodents. *Cell* 12: 1069-1084.
- Gall, J.G. and D.D. Atherton. 1974. Satellite DNA sequences in Drosophila virilis. *J. Mol. Biol.* 85: 633-664.
- Gall, J.G., E.H. Cohen and D.D. Atherton. 1974. The satellite DNAs of Drosophila virilis. *Cold Spring Harbor Symposium Quant. Biol.* 38: 417-421.
- Gall, J.G., E.H. Cohen and M.L. Polan. 1971. Repetitive DNA sequences in Drosophila. *Chromosoma* 33:319-344.
- Gilbert, W. 1978. Why genes in pieces? *Nature* 271: 501.
- Gilbert, W. and A. Maxam. 1973. The nucleotide sequence of the lac operator. *P.N.A.S.* 70: 3581-3584.
- Goodman, H.M. 1979. Repair of overlapping DNA termini. *Meth. Enzymol.* 65: 63-64.

- Gowen, J.W. and E.H. Gay. 1934. Chromosome constitution and behavior in eversporting and mottling in Drosophila melanogaster. Genetics 19: 189-208.
- Green, M.M. 1977. Genetic instability in Drosophila melanogaster: De novo induction of putative insertion mutations. P.N.A.S. 74: 3490-3493.
- Hartmann-Goldstein, I.J. 1967. On the relationship between heterochromatization and variegation in Drosophila, with special reference to temperature-sensitive periods. Genet. Res., Camb. 10: 143-159.
- Hennig, W. and P.M.B. Walker. 1970. Variations in the DNA from two rodent families (Cricetidae and Muridae). Nature 225: 915-919.
- Higgins, N.P. and N.R. Cozzarelli. 1979. DNA-joining enzymes: a review. Meth. Enzymol. 68: 50-71.
- Hill, R.J. and B.D. Stollar. 1983. Dependence of Z-DNA antibody binding to polytene chromosomes on acid fixation and DNA torsional strain. Nature 305: 338-340.
- Hourcade, D., D. Dressler and J. Wolfson. 1973. The amplification of ribosomal RNA genes involves a rolling circle intermediate. P.N.A.S. 70: 2926-2930.
- Hoyer, B.H., B.J. McCarthy and E.T. Bolton. 1964. A molecular approach in the systematics of higher organisms. Science 144: 959-967.
- Hsieh, T. and D.L. Brutlag. 1979a. A protein that preferentially binds Drosophila satellite DNA. P.N.A.S. 76: 726-730.



- Hsieh, T. and D. Brutlag. 1979b. Sequence and sequence variation within the 1.688 g/cm<sup>3</sup> satellite DNA of Drosophila melanogaster. J. Mol. Biol. 135: 465-481.
- Hurst, D.D., S. Fogel and R.K. Mortimer. 1972. Conversion - associated recombination in yeast. P.N.A.S. 69: 101-105.
- Iatrou, K., S.G. Tsitilou and F.C. Kafatos. 1984. DNA sequence transfer between two high-cysteine chorion gene families in the silkworm Bombyx mori. P.N.A.S. 81: 4452-4456.
- Iida, S., J. Meyer and W. Arber. 1983. Prokaryotic IS elements. Pp 159-221 in J.A. Shapiro, ed. Mobile Genetic Elements. Academic Press, New York.
- Ilyin, Y.V., N.A. Tchurikov, E.V. Ananiev, A.P. Ryskov, G.N. Yenikolopov, S.A. Limborska, N.E. Maleeva, V.A. Gvozdev and G.P. Georgiev. 1978. Studies on the DNA fragments of mammals and Drosophila containing structural genes and adjacent sequences. Cold Spring Harbor Symp. Quant. Biol. 42: 959-969.
- Jamrich, M., R. Warrior, R. Steele and J.G. Gall. 1983. Transcription of repetitive sequences on Xenopus lampbrush chromosomes. P.N.A.S. 80: 3364-3367.
- John, B. and M. King. 1985. The inter-relationship between heterochromatin distribution and chiasma distribution. Genetica, in press.
- John, B. and G.L.G. Miklos. 1979. Functional aspects of satellite DNA and heterochromatin. Int. Rev. Cytol. 58: 1-114.

- Jones, K.W. 1970. Chromosomal and nuclear location of mouse satellite DNA in individual cells. *Nature* 225: 912-915.
- Jones, K.W. and G. Corneo. 1971. Location of satellite and homogeneous DNA sequences on human chromosomes. *Nature New Biology* 233: 268-271.
- Kidwell, M.G., J.F. Kidwell and J.A. Sved. 1977. Hybrid dysgenesis in Drosophila melanogaster: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* 86: 813-833.
- Kit, S. 1961. Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. *J. Mol. Biol.* 3: 711-716.
- Kit, S. 1962a. Species differences in animal deoxyribonucleic acids as revealed by equilibrium sedimentation in density gradients. *Nature* 193: 274-275.
- Kit, S. 1962b. Physicochemical studies on the deoxyribonucleic acids of mouse tissues. Pp. 133-146 in R. Cumley, M. Abbott, D. Aldridge, J. Haroz, E. Heinsohn, J. McCay and B. Troller, eds. *The Molecular Basis of Neoplasia*. University of Texas Press, Austin.
- Klein, H.L. and T.D. Petes. 1981. Intrachromosomal gene conversion in yeast. *Nature*. 289:144-148.
- Kubinski, H., P. Gibbs and C.B. Kasper. 1972. Interactions of nucleic acids and cellular membranes: behavior of membranes in CsCl density gradients. *Biochim. Biophys. Acta* 281: 244-252.



- Kurnit, D.M., B.R. Shafit and J.J. Maio. 1973. Multiple satellite deoxyribonucleic acids in the calf and their relation to the sex chromosomes. *J. Mol. Biol.* 81: 273-284.
- Laird, C.B. and B.J. McCarthy. 1968. Magnitude of interspecific nucleotide sequence variability in Drosophila. *Genetics* 60: 303-322.
- Lam, B.S. and D. Carroll. 1983a. Tandemly repeated DNA sequences from Xenopus laevis. I. Studies on sequence organization and variation in satellite 1 DNA (741 base-pair repeat). *J. Mol. Biol.* 165: 567-585.
- Lam, B.S. and D. Carroll. 1983b. Tandemly repeated DNA sequences from Xenopus laevis. II. Dispersed clusters of a 388 base-pair repeating unit. *J. Mol. Biol.* 165: 587-597.
- Laskey, R.A. 1980. The use of intensifying screens or organic scintillators for visualizing radioactive molecules resolved by gel electrophoresis. *Meth. Enzymol.* 65: 363-371.
- Leblon, G. 1972. Mechanism of gene conversion in Ascobolus immersus I. Existence of a correlation between the origin of mutants induced by different mutagens and their conversion spectrum. *Molec. Gen. Genetics* 115: 36-48.
- Lederberg, E.M. and S.N. Cohen. 1974. Transformation of Salmonella typhimurium by plasmid deoxyribonucleic acid. *J. Bacteriol.* 119: 1072-1074.
- Lee, N.H.T. and M.F. Singer. 1982. Structural organization of  $\alpha$ -satellite DNA in a single monkey chromosome. *J. Mol. Biol.* 161: 323-342.

- Lemeunier, F. and M. Ashburner. 1976. Relationships within the melanogaster species subgroup of the genus Drosophila (Sophophora) II. Phylogenetic relationships between six species based upon polytene chromosome banding sequences. Proc. Roy. Soc. Lond. B. 193: 275-294.
- Levinger, L. and A. Varshavsky. 1982a. Selective arrangement of ubiquitinated and D1 protein - containing nucleosomes within the Drosophila genome. Cell 28: 375-385.
- Levinger, L. and A. Varshavsky. 1982b. Protein D1 preferentially binds A+T-rich DNA in vitro and is a component of Drosophila melanogaster nucleosomes containing A+T-rich satellite DNA. P.N.A.S. 79: 7152-7156.
- Lewis, E.B. 1950. The phenomenon of position effect. Adv. Genet. 3: 73-115.
- Lima-De-Faria, A. 1959. Differential uptake of tritiated thymidine into hetero- and euchromatin in Melanoplus and Secale. J. Biophys. Biochem. Cytol. 6: 457-466.
- Lima-De-Faria, A. and H. Jaworska. 1968. Late DNA synthesis in heterochromatin. Nature 217: 138-142.
- Maio, J.J. 1971. DNA strand reassociation and polyribonucleotide binding in the african green monkey, Cercopithecus aethiops. J. Mol. Biol. 56: 579-595.
- Manuelidis, L. and J.C. Wu. 1978. Homology between human and simian repeated DNA. Nature 276: 92-94.
- Maresca, A. and M.F. Singer. 1983. Deca-satellite: a highly polymorphic satellite that joins  $\alpha$ -satellite in the african green monkey genome. J. Mol. Biol. 164: 493-511.



- Maresca, A., M.F. Singer and T.N.H. Lee. 1984. Continuous reorganization leads to extensive polymorphism in a monkey centromeric satellite. *J. Mol. Biol.* 179: 629-649.
- Marmur, J., R. Rownd and C.L. Schildkraut. 1963. Denaturation and renaturation of deoxyribonucleic acid. *Prog. Nuc. Acid Res.* 1: 231-300.
- Maxam, A.M. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Meth. Enzymol.* 65: 499-559.
- Mayfield, J.E. and J.R. Ellison. 1975. The organization of interphase chromatin in *Drosophilidae*. The self adhesion of chromatin containing the same DNA sequences. *Chromosoma* 52: 37-48.
- Mazrimas, J.A. and F.T. Hatch. 1972. A possible relationship between satellite DNA and the evolution of kangaroo rat species (Genus Dipodomys). *Nature New Biol.* 240: 102-105.
- McCutchan, T., H. Hsu, R.E. Thayer and M.F. Singer. 1982. Organization of african green monkey DNA at junctions between  $\alpha$ -satellite and other DNA sequences. *J. Mol. Biol.* 157: 195-211.
- Messing, J. 1981. M13 mp2 and derivatives: a molecular cloning system for DNA sequencing, strand specific hybridization, and in vitro mutagenesis. Pp 143-153 in A.G. Walton, ed. *Proceedings of the third Cleveland symposium on macromolecules.* Elsevier, Amsterdam.
- Messing, J. and J. Vierira. 1982. A new pair of M13 vectors for selecting either strand of double-digest restriction fragments. *Gene* 19: 269-276.

- Meyerhof, W., B. Tappeser, E. Korge and W. Knochel. 1983. Satellite DNA from Xenopus laevis: comparative analysis of 745 and 1037 base pair Hind III tandem repeats. N.A.R. 11: 6997-7009.
- Miklos, G.L.G. 1985. Localised highly repetitive DNA sequences in vertebrate and invertebrate genomes. *Evol. Theory.*, in press.
- Miklos, G.L.G. and A.C. Gill. 1981. The DNA sequences of cloned complex satellite DNAs from hawaiian Drosophila and their bearing on satellite DNA sequence conservation. *Chromosoma* 82: 409-427.
- Miklos, G.L.G. and R.N. Nankivell. 1976. Telomeric satellite DNA functions in regulating recombination. *Chromosoma* 56: 143-167.
- Moran, C. 1979. The structure of the hybrid zone in Caledia captiva. *Heredity* 42: 13-32.
- Moran, C. and D.D. Shaw. 1977. Population cytogenetics of the genus Caledia (Orthoptera: Acridinae) III. Chromosomal polymorphisms, racial parapatry and introgression. *Chromosoma* 63: 181-204.
- Moran, C., P. Wilkinson and D.D. Shaw. 1980. Allozyme variation across a narrow hybrid zone in the grasshopper Caledia captiva. *Heredity* 44: 69-81.
- Moss, T. 1983. A transcriptional function for the repetitive ribosomal spacer in Xenopus laevis. *Nature* 302: 223-228.



- Moyzis, R.K., J. Bonnet, D.W. Li and P.O.P. Ts'o. 1981. An alternative view of mammalian DNA sequence organization I. Repetitive sequence interspersions in Syrian hamster DNA: a model system. *J. Mol. Biol.* 153: 841-870.
- Muller, H.J. 1930. Types of visible variations induced by X-rays in Drosophila. *J. Genet.* 22: 299-334.
- Muller, F., P. Walker, P. Aeby, H. Neuhaus, H. Felder, E. Back and H. Tobler. 1982. Nucleotide sequence of satellite DNA contained in the eliminated genome of Ascaris lumbricoides. *N.A.R.* 10: 7493-7510.
- Nagl, W. 1974. Role of heterochromatin in the control of cell cycle duration. *Nature* 249: 53-54.
- Nei, M. 1972. Genetic distance between populations. *Amer. Nature* 106: 283-292.
- Nei, M. and A.K. Roychoudhury. 1974. Sampling variances of heterozygosity and genetic distance. *Genetics* 76: 379-390.
- Newport, J. and M. Kirschner. 1982a. A major developmental transition in early Xenopus embryos: I. Characterization and timing of cellular changes at the midblastula stage. *Cell* 30: 675-686.
- Newport, J. and M. Kirschner. 1982b. A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. *Cell* 30: 687-696.
- Nordheim, A., M.L. Pardue, E.M. Lafer, A. Moller, B.D. Stollar and A. Rich. 1981. Antibodies to left-handed Z-DNA bind to interband regions of Drosophila polytene chromosomes. *Nature* 294: 417-422.

- Novak, U. 1984. Structure and properties of a highly repetitive DNA sequence in sheep. NAR 12: 2343-2350.
- Ohta, T. and G.A. Dover. 1983. Population genetics of multigene families that are dispersed into two or more chromosomes. P.N.A.S. 80: 4079-4083.
- Ohta, T. and G.A. Dover. 1984. The cohesive population genetics of molecular drive. Genetics 108: 501-521.
- Orgel, L.E. and F.H.C. Crick. 1980. Selfish DNA: the ultimate parasite. Nature 284: 604-607.
- Pardue, M.L. and J.G. Gall. 1969. Molecular hybridization of radioactive DNA to the DNA of cytological preparations. P.N.A.S. 64: 600-604.
- Pardue, M.L. and J.G. Gall. 1970. Chromosomal localization of mouse satellite DNA. Science 168: 1356-1358.
- Pardue, M.L., S.A. Gerbi, R.A. Eckhardt and J.G. Gall. 1970. Cytological localization of DNA complementary to ribosomal RNA in polytene chromosomes of Diptera. Chromosoma 29: 268-290.
- Parker, D.R. 1967. Induced heterologous exchange at meiosis in Drosophila I. Exchanges between Y and fourth chromosomes. Mutation Res. 4: 333-337.
- Peacock, W.J., R. Appels, P. Dunsmuir, A.R. Lohe and W.L. Gerlach. 1977. Highly repeated DNA sequences: chromosomal localization and evolutionary conservatism. Pp 494-506 in B.R. Brinkley and K.R. Porter, eds. International Cell Biology. Rockefeller University Press, New York.



- Peacock, W.J., D. Brutlag, E. Goldring, R. Appels, C.W. Hinton and D.L. Lindsley. 1974. The organization of highly repeated DNA sequences in Drosophila melanogaster chromosomes. Cold Spring Harbor Symp. Quant. Biol. 38: 405-421.
- Peacock, W.J., A.R. Lohe, W.L. Gerlach, P. Dunsmuir, E.S. Dennis and R. Appels. 1978. Fine structure and evolution of DNA in heterochromatin. Cold Spring Harbor Symp. Quant. Biol. 42: 1121-1135.
- Petes, T.D. 1980. Unequal meiotic recombination within tandem arrays of yeast ribosomal DNA genes. Cell 19: 765-774.
- Pietras, D.F., K.L. Bennett, L.D. Siracusa, M. Woodworth-Gutai, V.M. Chapman, K.W. Gross, C. Kane-Haas and N.D. Hastie. 1983. Construction of a small Mus musculus repetitive DNA library: identification of a new satellite sequence in Mus musculus. N.A.R. 11: 6965-6983.
- Plucienniczak, A., J. Skowronski and J. Jaworski. 1982. Nucleotide sequence of bovine 1.715 satellite DNA and its relation to other bovine satellite sequences. J. Mol. Biol. 158: 293-304.
- Potter, S.S., W.J. Brorein, Jr., P. Dunsmuir and G.M. Rubin. 1979. Transposition of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila. Cell 17: 415-427.
- Ptashne, M., A. Jeffrey, A.D. Johnson, R. Maurer, B.J. Meyer, C.O. Pabo, T.M. Roberts and R.T. Sauer. 1980. How the  $\lambda$  repressor and Cro work. Cell 19: 1-11.

Rees, H., D.D. Shaw and P. Wilkinson. 1978. Nuclear DNA  
Variation among acridid grasshoppers. Proc. R. Soc.  
Lond. B. 202: 517-525.



- Rae, P.M.M. 1970. Chromosomal distribution of rapidly reannealing DNA in Drosophila melanogaster. P.N.A.S. 67: 1018-1025.
- Rae, P.M.M. and W.W. Franke. 1972. The interphase distribution of satellite DNA-containing heterochromatin in mouse nuclei. Chromosoma 39: 443-456.
- Ratty, F.J. 1954. Gene action and position effect in duplications in Drosophila melanogaster. Genetics 39: 513-528.
- Rigby, W.J.P., M. Dieckmann, C. Rhodes and P. Berg. 1977. Labelling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. 113: 237-251.
- Robert-Nicoud, M., D.J. Arndt-Jovin, D.A. Zarling and T.M. Jovin. 1984. Immunological detection of left-handed Z-DNA in isolated polytene chromosomes. Effects of ionic strength, pH, temperature and topological stress. EMBO J. 3: 721-731.
- Roberts, R.J. 1979. Directory of restriction endonucleases. Method. Enzymol. 68: 27-41.
- Rolfe, R. and M. Meselson. 1959. The relative homogeneity of microbial DNA. P.N.A.S. 45: 1039-1045.
- Roman, H. 1963. Genic conversion in Fungi. Pp. 209-227 in W.J. Burdette, ed. Methodology in Basic Genetics. Holden-Day, Inc., San Francisco.
- Rosenberg, H., M. Singer and M. Rosenberg. 1978. Highly reiterated sequences of SIMIANSIMIANSIMIANSIMIAN. Science 200: 394-402.

- Rubin, C.M., P.L. Deininger, C.M. Houck and C.W. Schmid. 1980.  
A dimer satellite sequence in Bonnet monkey DNA consists of  
distinct monomer subunits. J. Mol. Biol. 136: 151-167.
- Rubin, G.M. 1983. Dispersed repetitive DNAs in Drosophila. Pp.  
329-361 in J.A. Shapiro, ed. Mobile genetic elements.  
Academic Press, New York.
- Rushlow, C.A., W. Bender and A. Chovnick. 1984. Studies on the  
mechanism of heterochromatic position effect at the rosy  
locus of Drosophila melanogaster. Genetics 108: 603-615.
- Sanger, F., S. Nicklen and A.R. Coulson. 1977. DNA sequencing  
with chain-terminating inhibitors. P.N.A.S. 74: 5463-5467.
- Shaw, D.D. 1976. Population cytogenetics of the genus Caledia  
(Orthoptera: Acridinae) I. Inter- and intraspecific  
karyotype diversity. Chromosoma 54: 221-243.
- Shaw, D.D. and D.J. Coates. 1983. Chromosomal variation and the  
concept of the coadapted genome - a direct cytological  
assessment. Pp. 207-216 in P.E. Brandham and M.D. Bennett,  
eds. Kew Chromosome Conference II. G. Allen and Unwin,  
London.
- Shaw, D.D., D.J. Coates, M.L. Arnold and P. Wilkinson. 1985.  
Temporal variation in the chromosomal structure of a hybrid  
zone and its relationship to karyotypic repatterning.  
Heredity, in press.
- Shaw, D.D. and G.R. Knowles. 1976. Comparative chiasma analysis  
using a computerised optical digitiser. Chromosoma 59:  
103-127.



- Shaw, D.D., C. Moran and P. Wilkinson. 1980. Chromosomal reorganization, geographic differentiation and the mechanism of speciation in the genus Caledia. Pp. 171-194 in R.L. Blackman, G.M. Hewitt and M. Ashburner, eds. Insect Cytogenetics, Symp. Roy. Entomol. Soc. London No 10. Blackwells, Oxford.
- Shaw, D.D., G.C. Webb and P. Wilkinson. 1976. Population cytogenetics of the genus Caledia (Orthoptera: Acridinae) II. Variation in the pattern of C-banding. Chromosoma 56: 169-190.
- Shaw, D.D. and P. Wilkinson. 1978. "Homologies" between non-homologous chromosomes in the grasshopper Caledia captiva. Chromosoma 68: 241-259.
- Shaw, D.D. and P. Wilkinson. 1980. Chromosome differentiation, hybrid breakdown and the maintenance of a narrow hybrid zone in Caledia. Chromosoma 80: 1-31.
- Shaw, D.D., P. Wilkinson and D.J. Coates. 1982. The chromosomal component of reproductive isolation in the grasshopper Caledia captiva. II. The relative viabilities of recombinant and non-recombinant chromosomes during embryogenesis. Chromosoma 86: 533-549.
- Shaw, D.D., P. Wilkinson and C. Moran. 1979. A comparison of chromosomal and allozymal variation across a narrow hybrid zone in the grasshopper Caledia captiva. Chromosoma 75: 333-351.
- Shmookler Reis, R.J. and P.A. Biro. 1978. Sequence and evolution of mouse satellite DNA. J. Mol. Biol. 121: 357-374.

- Siebenlist, U., R.B. Simpson and W. Gilbert. 1980. E. coli RNA polymerase interacts homologously with two different promoters. Cell 20: 269-281.
- Singer, M.F. 1982. Highly repeated sequences in mammalian genomes. Int. Rev. Cytol. 76: 67-112.
- Skinner, D.M. and M.S. Kerr. 1971. Characterization of mitochondrial and nuclear satellite deoxyribonucleic acids of five species of Crustacea. Biochemistry 10: 1864-1872.
- Slightom, J.L., A.E. Blechl and O. Smithies. 1980. Human fetal G $\gamma$ - and A $\gamma$ - globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell 21: 627-638.
- Smith, D.B. and R.B. Flavell. 1975. Characterisation of the wheat genome by renaturation kinetics. Chromosoma 50: 223-242.
- Smith, G.P. 1976. Evolution of repeated DNA sequences by unequal crossover. Science 191: 528-535.
- Southern, E.M. 1970. Base sequence and evolution of guinea-pig  $\alpha$ -satellite DNA. Nature 227: 794-798.
- Southern, E.M. 1975a. Long range periodicities in mouse satellite DNA. J. Mol. Biol. 94: 51-69.
- Southern, E.M. 1975b. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98: 503-517.



- Spofford, J.B. 1976. Position-effect variegation in Drosophila. Pp. 955-1018 in M. Ashburner and E. Novitski, eds. The genetics and biology of Drosophila. 1c. Academic Press, New York.
- Strachan, T., E. Coen, D. Webb and G. Dover. 1982. Modes and rates of change of complex DNA families of Drosophila. J. Mol. Biol. 158: 37-54.
- Strachan, T., D. Webb and G. Dover. 1985. Transition stages during sequence turnover in multiple-copy DNA families in Drosophila. Cell, in press.
- Strauss, F. and A. Varshavsky. 1984. A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. Cell 37: 889-901.
- Strobel, E., P. Dunsmuir and G.M. Rubin. 1979. Polymorphisms in the chromosomal locations of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila. Cell 17: 429-439.
- Sueoka, N. 1961. Variation and heterogeneity of base composition of deoxyribonucleic acids: a compilation of old and new data. J. Mol. Biol. 3: 31-40.
- Sueoka, N. and T. Cheng. 1962. Natural occurrence of deoxyribonucleic acid resembling the deoxyadenylate - deoxythymidylate polymer. P.N.A.S. 48: 1851-1856.
- Sutton, W.D. and M. McCallum. 1972. Related satellite DNA's in the genus Mus. J. Mol. Biol. 71: 633-656.

- Szostak, J.W. and R. Wu. 1980. Unequal crossing over in the ribosomal DNA of Saccaromyces cerevisiae. Nature 284: 426-430.
- Szybalski, W. 1962. Properties and applications of halogenated deoxyribonucleic acids. Pp 147-171 in R. Cumley, M. Abbott, D. Aldridge, J. Haroz, E. Heinsohn, J. McCay and B. Troller, eds. The Molecular Basis of Neoplasia. Univ. of Texas Press, Austin.
- Taparowsky, E.J. and S.A. Gerbi. 1982. Sequence analysis of bovine satellite I DNA (1.715 gm/cm<sup>3</sup>). N.A.R. 10: 1271-1281.
- Tartof, K.D., C. Hobbs and M. Jones. 1984. A structural basis for variegating position effects. Cell 37: 869-878.
- Thayer, R.E., M.F. Singer and T.F. McCutchan. 1981. Sequence relationships between single repeat units of highly reiterated African Green monkey DNA. N.A.R. 9: 169-181.
- Trick, M. and G.A. Dover. 1984. Unexpectedly slow homogenisation within a repetitive DNA family shared between two subspecies of tsetse fly. J. Mol. Evol. 20: 322-329.
- Ullrich, A., J. Shine, J. Chirgwin, R. Pictet, E. Tischer, W.J. Rutter and H.M. Goodman. 1977. Rat insulin genes: construction of plasmids containing the coding sequences. Science 196: 1313-1319.
- Walker, P.M.B. 1971. Origin of satellite DNA. Nature 229: 306-308.
- Walker, P.M.B. and A. McLaren. 1965. Fractionation of mouse deoxyribonucleic acid on hydroxyapatite. Nature 208: 1175-1179.



- Wang, A. H.-J., G.J. Quigley, F.J. Kolpak, J.L. Crawford, J.H. van Boom, G. Vander Marel and A. Rich. 1979. Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* 282: 680-686.
- Waring, M. and R.J. Britten. 1966. Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. *Science* 154: 791-794.
- Watson, J.D. and F.H.C. Crick. 1953. Molecular structure of nucleic acids. *Nature* 171: 737-738.
- Webb, G.C. 1976. Chromosome organisation in the Australian plague locust, Chortoicetes terminifera I. Banding relationships of the normal and supernumerary chromosomes. *Chromosoma* 55: 229-246.
- Witney, F.R. and A.V. Furano. 1983. The independent evolution of two closely related satellite DNA elements in rats (Rattus). *N.A.R.* 11: 291-304.
- Yu-Sun, C.C., M.R.T. Wickramaratne and H.L.K. Whitehouse. 1977. Mutagen specificity in conversion pattern in Sordaria brevicollis. *Genet. Res.* 29: 65-81.
- Zimmer, E.A., S.L. Martin, S.M. Beverley, Y.W. Kan and A.C. Wilson. 1980. Rapid duplication and loss of genes coding for the  $\alpha$  chains of hemoglobin. *P.N.A.S.* 77: 2158-2162.